

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number  
**WO 02/097059 A2**

(51) International Patent Classification<sup>7</sup>: C12N

(21) International Application Number: PCT/US02/17452

(22) International Filing Date: 30 May 2002 (30.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/294,758 30 May 2001 (30.05.2001) US  
60/366,891 21 March 2002 (21.03.2002) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US 60/294,758 (CIP)  
Filed on 30 May 2001 (30.05.2001)  
US 60/366,891 (CIP)  
Filed on 21 March 2002 (21.03.2002)

(71) Applicant (for all designated States except US): CHROMOS MOLECULAR SYSTEMS, INC. [CA/CA]; 8081 Lougheed Highway, Burnaby, B.C. V5A 1W9 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): PERKINS, Edward [US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6 (CA). PEREZ, Carl [US/CA]; 1201-7680 Granville Avenue, Richmond, B.C. V6Y 4B9 (CA). LINDENBAUM, Michael [CA/CA]; 252 Finnigan Street, Coquitlam, B.C. V3K 5J7 (CA). GREENE, Amy [US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6 (CA). LEUNG, Josephine [CA/CA]; 711 Ebert Avenue, Coquitlam, B.C. V3J 7P8 (CA). FLEMING, Elena [CA/CA]; 248 E 18th, North Vancouver, B.C. V7L 2X6 (CA). STEWART, Sandra [CA/CA]; 2618 Oxford Street, Vancouver, B.C. V5K 1N3 (CA). SHELLARD, Joan [CA/CA]; #215-1345 West 15th Avenue, Vancouver, B.C. V6H 3R3 (CA).

(74) Agents: SEIDMAN, Stephanie, L. et al.; Heller Ehrman White & McAuliffe LLP, 7th floor, 4350 La Jolla Village Drive, San Diego, CA 92122-1246 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,

MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Declarations under Rule 4.17:**

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: CHROMOSOME-BASED PLATFORMS

(57) Abstract: Artificial chromosomes, including *Aces*, that have been engineered to contain available sites for site-specific, recombination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational engineering of the chromosome for a variety of applications.

WO 02/097059 A2

1-

**CHROMOSOME-BASED PLATFORMS****RELATED APPLICATIONS**

Benefit of priority to U.S. provisional application Serial No. 60/294,758, filed May 30, 2001, to Perkins, *et al.*, entitled "CHROMOSOME-BASED PLATFORMS" and to U.S. provisional application Serial No. 60/366,891, filed March 21, 2002, to Perkins, *et al.*, entitled

5 "CHROMOSOME-BASED PLATFORMS" is claimed. Where permitted, the subject matter of which are herein incorporated by reference in their entirety.

This application is related to Provisional Application No. 60/294,687, filed May 30, 2001, by CARL PEREZ AND STEVEN

10 FABIJANSKI entitled *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING PLANT ARTIFICIAL CHROMOSOMES* and to U.S. Provisional Application No. 60/296,329, filed June 4, 2001, by CARL PEREZ AND STEVEN FABIJANSKI entitled

15 *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING PLANT ARTIFICIAL CHROMOSOMES*. This application also is related to U.S. Provisional Application No. 60/294,758, filed May 30, 2001, by EDWARD PERKINS *et al.*, entitled *CHROMOSOME-BASED PLATFORMS* and to U.S. Provisional Application No. 60/366,891, filed March 21, 2002, by by EDWARD PERKINS *et al.*, entitled

20 *CHROMOSOME-BASED PLATFORMS*. This application is also related to U.S. application Serial Nos. (attorney dkt nos. 24601-419 and 419PC), filed on the same day herewith, entitled *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS OF PREPARING PLANT ARTIFICIAL CHROMOSOMES* to Perez *et al.*.

25 . This application is related to U.S. application Serial No. 08/695,191, filed August 7, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND*



-2-

*METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*, now U.S. Patent No. 6,025,155. This application is also related to U.S. application Serial No. 08/682,080, filed July 15, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF*

5 *AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*, now U.S. Patent No. 6,077,697. This application is also related U.S. application Serial No. 08/629,822, filed April 10, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL*

10 *CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* (now abandoned), and is also related to copending U.S. application Serial No. 09/096,648, filed June 12, 1998, by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING*

15 *ARTIFICIAL CHROMOSOMES* and to U.S. application Serial No. 09/835,682, April 10, 1997 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* (now abandoned). This application is also related to copending U.S. application Serial

20 No. 09/724,726, filed November 28, 2000, U.S. application Serial No. 09/724,693, filed November 28, 2000, U.S. application Serial No. 09/799,462, filed March 5, 2001, U.S. application Serial No. 09/836,911, filed April 17, 2001, and U.S. application Serial No. 10/125,767, filed April 17, 2002, each of which is by GYULA

25 HADLACZKY and ALADAR SZALAY, and is entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*. This application is also related to International PCT application No. WO 97/40183. Where permitted the

-3-

subject matter of each of these provisional applications, international applications, and applications is incorporated by reference in its entirety.

#### **FIELD OF INVENTION**

Artificial chromosomes, including *ACes*, that have been engineered  
5 to contain available sites for site-specific, recombination-directed  
integration of DNA of interest are provided. These artificial chromosomes  
permit tractable, efficient, rational engineering of the chromosome.

#### **BACKGROUND**

##### **Artificial chromosomes**

10 A variety of artificial chromosomes for use in plants and animals,  
particularly higher plants and animals are available. In particular, U.S.  
Patent Nos. 6,025,155 and 6,077,697 provide heterochromatic artificial  
chromosomes designated therein as satellite artificial chromosomes  
15 (*SATACs*) and now designated artificial chromosome expression systems  
(*ACes*). These chromosomes are prepared by introducing heterologous  
DNA into a selected plant or animal cell under conditions that result in  
integration into a region of the chromosome that leads to an amplification  
event resulting in production of a dicentric chromosome. Subsequent  
treatment and growth of cells with dicentric chromosomes, including  
20 further amplifications, ultimately results in the artificial chromosomes  
provided therein. In order to introduce a desired heterologous gene (or a  
plurality of heterologous genes) into the artificial chromosome, the  
process is repeated introducing the desired heterologous genes and  
nucleic acids in the initial targeting step. This process is time consuming  
25 and tedious. Hence, more tractable and efficient methods for introducing  
heterologous nucleic acid molecules into artificial chromosomes,  
particularly *ACes*, are needed.

-4-

Therefore, it is an object herein to provide engineered artificial chromosomes that permit tractable, efficient and rational engineering of artificial chromosomes.

#### **SUMMARY OF THE INVENTION**

5            Provided herein are artificial chromosomes that permit tractable, efficient and rational engineering thereof. In particular, the artificial chromosomes provided herein contain one or a plurality of loci (sites) for site-specific, recombination-directed integration of DNA. Thus, provided herein are platform artificial chromosome expression systems ("platform  
10 *ACes*") containing single or multiple site-specific, recombination sites. The artificial chromosomes and *ACes* artificial chromosomes include plant and animal chromosomes. Any recombinase system that effects site-specific recombination is contemplated for use herein.

          In one embodiment, chromosomes, including platform *ACes*, are  
15 provided that contain one or more lambda *att* sites designed for recombination-directed integration in the presence of lambda integrase, and that are mutated so that they do not require additional factors. Methods for preparing such chromosomes, vectors for use in the methods, and uses of the resulting chromosomes are also provided.

20            Platform *ACes* containing the recombination site(s) and methods for introducing heterologous nucleic acid into such sites and vectors therefor, are provided.

          Also provided herein is a bacteriophage lambda ( $\lambda$ ) integrase site-specific recombination system.

25            Methods using recombinase mediated recombination target gene expression vectors and/or genes for insertion thereof into platform chromosomes and the resulting chromosomes are provided.

          Combinations and kits containing the combinations of vectors encoding a recombinase and integrase and primers for introduction of the

-5-

site recognized thereby are also provided. The kits optionally include instructions for performing site-directed integration or preparation of *ACes* containing such sites.

Also provided herein are mammalian and plant cells comprising the  
5 artificial chromosomes and *ACes* described herein. The cells can be nuclear donor cells, stem cells, such as a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell or an embryonic stem cell.

Also provide is a lambda-intR mutain comprising a glutamic acid to arginine change at position 174 of wild-type lambda-integrase3. Also  
10 provided are transgenic animals and methods for producing a transgenic animal, comprising introducing a *ACes* into an embryonic cell, such as a stem cell or embryo. The *ACes* can comprise heterologous nucleic acid that encodes a therapeutic product. The transgenic animal can be a fish, insect, reptile, amphibians, arachnid or mammal. In certain embodiments,  
15 the *ACes* is introduced by cell fusion, lipid-mediated transfection by a carrier system, microinjection, microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer.

The platform *ACes*, including plant and animal *ACes*, such as MACs, provided herein can be introduced into cells, such as, but not  
20 limited to, animal cells, including mammalian cells, and into plant cells. Hence plant cells that contain platform MACs, animal cells that contain platform PACs and other combinations of cells and platform *ACes* are provided.

#### DESCRIPTION OF FIGURES

25 FIGURE 1 provides a diagram depicting creation of an exemplary *ACes* artificial chromosome prepared using methods detailed in U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183. In this exemplified embodiment, the nucleic acid is targeted to an acrocentric chromosome in an animal or plant, and the

-6-

heterologous nucleic acid includes a sequence-specific recombination site and marker genes.

FIGURE 2 provides a map of pWEPuro9K, which is a targeting vector derived from the vector pWE15 (GenBank Accession # X65279; SEQ ID No. 31). Plasmid pWE15 was modified by replacing the *Sal*I (Klenow filled)/*Sma*I neomycin resistance encoding fragment with the *Pvu*II/*Bam*HI (Klenow filled) puromycin resistance-encoding fragment (isolated from plasmid pPUR, Clontech Laboratories, Inc., Palo Alto, CA; GenBank Accession no. U07648; SEQ ID No. 30) resulting in plasmid pWEPuro. Subsequently a 9 Kb *Not*I fragment from the plasmid pFK161 (see Example 1, see, also Csonka *et al.* (2000) *Journal of Cell Science* 113:3207-32161; and SEQ ID NO: 118), containing a portion of the mouse rDNA region, was cloned into the *Not*I site of pWEPuro resulting in plasmid pWEPuro9K.

FIGURE 3 depicts construction of an *ACes* platform chromosome with a single recombination site, such as loxP sites or an *att*P or *att*B site. This platform *ACes* chromosome is an exemplary artificial chromosome with a single recombination site.

FIGURE 4 provides a map of plasmid pSV40-193attPsensePur.

FIGURE 5 depicts a method for formation of a chromosome platform with multiple recombination integration sites, such as *att*P sites.

FIGURE 6 sets forth the sequences of the core region of *att*P, *att*B, *att*L and *att*R (SEQ ID Nos. 33-36).

FIGURE 7 depicts insertional recombination of a vector encoding a marker gene, DsRed and an *att*B site with an artificial chromosome containing an *att*P site.

FIGURE 8 provides a map of plasmid pCXLamIntR (SEQ ID NO: 112), which includes the Lambda integrase (E174R)-encoding nucleic acid.

-7-

FIGURE 9 diagrammatically summarizes the platform technology; marker 1 permits selection of the artificial chromosomes containing the integration site; marker 2, which is promoterless in the target gene expression vector, permits selection of recombinants. Upon  
5 recombination with the platform marker 2 is expressed under the control of a promoter resident on the platform.

FIGURE 10 provides the vector map for the plasmid p18attBZEO-5'6XHS4eGFP (SEQ ID NO: 116).

FIGURE 11 provides the vector map for the plasmid p18attBZEO-  
10 3'6XHS4eGFP (SEQ ID NO: 115).

FIGURE 12 provides the vector map for the plasmid p18attBZEO-(6XHS4)2eGFP (SEQ ID NO: 110).

FIGURES 13 AND 14 depict the integration of a PCR product by site-specific recombination as set forth in Example 8.

FIGURE 15 provides the vector map for the plasmid pPACrDNA as  
15 set forth in Example 9.A.

## **DETAILED DESCRIPTION OF THE INVENTION**

### **A. DEFINITIONS**

Unless defined otherwise, all technical and scientific terms used  
20 herein have the same meaning as is commonly understood by one of skill in the art to which the invention(s) belong. All patents, patent applications, published applications and publications, Genbank sequences, websites and other published materials referred to throughout the entire disclosure herein, unless noted otherwise, are incorporated by reference  
25 in their entirety. Where reference is made to a URL or other such identifier or address, it understood that such identifiers can change and particular information on the internet can come and go, but equivalent information can be found by searching the internet. Reference thereto evidences the availability and public dissemination of such information.

-8-

As used herein, nucleic acid refers to single-stranded and/or double-stranded polynucleotides, such as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), as well as analogs or derivatives of either RNA or DNA. Also included in the term "nucleic acid" are analogs of nucleic acids such as peptide nucleic acid (PNA), phosphorothioate DNA, and other such analogs and derivatives. When referring to probes or primers, optionally labeled, with a detectable label, such as a fluorescent or radiolabel, single-stranded molecules are contemplated. Such molecules are typically of a length such that they are statistically unique and of low copy number (typically less than 5, preferably less than 3) for probing or priming a library. Generally a probe or primer contains at least 14, 16 or 30 contiguous nucleotides of sequence complementary to or identical to a gene of interest. Probes and primers can be 10, 20, 30, 50, 100 or more nucleotides long.

As used herein, DNA is meant to include all types and sizes of DNA molecules including cDNA, plasmids and DNA including modified nucleotides and nucleotide analogs.

As used herein, nucleotides include nucleoside mono-, di-, and triphosphates. Nucleotides also include modified-nucleotides, such as, but are not limited to, phosphorothioate nucleotides and deazapurine nucleotides and other nucleotide analogs.

As used herein, heterologous or foreign DNA and RNA are used interchangeably and refer to DNA or RNA that does not occur naturally as part of the genome in which it is present or which is found in a location or locations and/or in amounts in a genome or cell that differ from that in which it occurs in nature. Heterologous nucleic acid is generally not endogenous to the cell into which it is introduced, but has been obtained from another cell or prepared synthetically. Generally, although not necessarily, such nucleic acid encodes RNA and proteins that are not

-9-

normally produced by the cell in which it is expressed. Any DNA or RNA that one of skill in the art would recognize or consider as heterologous or foreign to the cell in which it is expressed is herein encompassed by heterologous DNA. Heterologous DNA and RNA may also encode RNA or proteins that mediate or alter expression of endogenous DNA by affecting transcription, translation, or other regulatable biochemical processes.

Examples of heterologous DNA include, but are not limited to, DNA that encodes a gene product or gene product(s) of interest, introduced for purposes of modification of the endogenous genes or for production of an encoded protein. For example, a heterologous or foreign gene may be isolated from a different species than that of the host genome, or alternatively, may be isolated from the host genome but operably linked to one or more regulatory regions which differ from those found in the unaltered, native gene. Other examples of heterologous DNA include, but are not limited to, DNA that encodes traceable marker proteins, such as a protein that confers traits including, but not limited to, herbicide, insect, or disease resistance; traits, including, but not limited to, oil quality or carbohydrate composition. Antibodies that are encoded by heterologous DNA may be secreted or expressed on the surface of the cell in which the heterologous DNA has been introduced.

As used herein, operative linkage or operative association, or grammatical variations thereof, of heterologous DNA to regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences refers to the relationship between such DNA and such sequences of nucleotides. For example, operative linkage of heterologous DNA to a promoter refers to the physical relationship between the DNA and the promoter such that the transcription of such DNA is initiated from the



-10-

promoter by an RNA polymerase that specifically recognizes, binds to and transcribes the DNA.

In order to optimize expression and/or *in vitro* transcription, it may be necessary to remove, add or alter 5' untranslated portions of the clones to eliminate extra, potential inappropriate alternative translation initiation (*i.e.*, start) codons or other sequences that may interfere with or reduce expression, either at the level of transcription or translation. Alternatively, consensus ribosome binding sites (see, *e.g.*, Kozak (1991) *J. Biol. Chem.* 266:19867-19870) can be inserted immediately 5' of the start codon and may enhance expression.

As used herein, a sequence complementary to at least a portion of an RNA, with reference to antisense oligonucleotides, means a sequence having sufficient complementarity to be able to hybridize with the RNA, preferably under moderate or high stringency conditions, forming a stable duplex. The ability to hybridize depends on the degree of complementarity and the length of the antisense nucleic acid. The longer the hybridizing nucleic acid, the more base mismatches it can contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

As used herein, regulatory molecule refers to a polymer of deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) or a polypeptide that is capable of enhancing or inhibiting expression of a gene.

As used herein, recognition sequences are particular sequences of nucleotides that a protein, DNA, or RNA molecule, or combinations thereof, (such as, but not limited to, a restriction endonuclease, a modification methylase and a recombinase) recognizes and binds. For example, a recognition sequence for Cre recombinase (see, *e.g.*, SEQ ID NO:58) is a 34 base pair sequence containing two 13 base pair inverted

-11-

repeats (serving as the recombinase binding sites) flanking an 8 base pair core and designated loxP (see, *e.g.*, Sauer (1994) *Current Opinion in Biotechnology* 5:521-527). Other examples of recognition sequences, include, but are not limited to, attB and attP, attR and attL and others  
5 (see, *e.g.*, SEQ ID Nos. 8, 41-56 and 72), that are recognized by the recombinase enzyme Integrase (see, SEQ ID Nos. 37 and 38 for the nucleotide and encoded amino acid sequences of an exemplary lambda phage integrase).

The recombination site designated attB is an approximately 33 base  
10 pair sequence containing two 9 base pair core-type Int binding sites and a 7 base pair overlap region; attP (SEQ ID No. 72) is an approximately 240 base pair sequence containing core-type Int binding sites and arm-type Int binding sites as well as sites for auxiliary proteins IHF, FIS, and Xis (see, *e.g.*, Landy (1993) *Current Opinion in Biotechnology* 3:699-707) see,  
15 *e.g.*, SEQ ID Nos. 8 and 72).

As used herein, a recombinase is an enzyme that catalyzes the exchange of DNA segments at specific recombination sites. An integrase herein refers to a recombinase that is a member of the lambda ( $\lambda$ ) integrase family.

20 As used herein, recombination proteins include excisive proteins, integrative proteins, enzymes, co-factors and associated proteins that are involved in recombination reactions using one or more recombination sites (see, Landy (1993) *Current Opinion in Biotechnology* 3:699-707). The recombination proteins used herein can be delivered to a cell via an  
25 expression cassette on an appropriate vector, such as a plasmid, and the like. In other embodiments, the recombination proteins can be delivered to a cell in protein form in the same reaction mixture used to deliver the desired nucleic acid, such as a platform ACes, donor target vectors, and the like.

-12-

As used herein the expression "lox site" means a sequence of nucleotides at which the gene product of the cre gene, referred to herein as Cre, can catalyze a site-specific recombination event. A LoxP site is a 34 base pair nucleotide sequence from bacteriophage P1 (see, 5 *e.g.*, Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398-3402). The LoxP site contains two 13 base pair inverted repeats separated by an 8 base pair spacer region as follows: (SEQ ID NO. 57):

ATAACTTCGTATA ATGTATGC TATACGAAGTTAT

*E. coli* DH5Δlac and yeast strain BSY23 transformed with plasmid pBS44 10 carrying two loxP sites connected with a LEU2 gene are available from the American Type Culture Collection (ATCC) under accession numbers ATCC 53254 and ATCC 20773, respectively. The lox sites can be isolated from plasmid pBS44 with restriction enzymes *EcoRI* and *Sall*, or 15 *XhoI* and *BamHI*. In addition, a preselected DNA segment can be inserted into pBS44 at either the *Sall* or *BamHI* restriction enzyme sites. Other lox sites include, but are not limited to, LoxB, LoxL, LoxC2 and LoxR sites, which are nucleotide sequences isolated from *E. coli* (see, *e.g.*, Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398). Lox sites can also be produced by a variety of synthetic techniques (see, *e.g.*, Ito *et al.* (1982) 20 *Nuc. Acid Res.* 10:1755 and Ogilvie *et al.* (1981) *Science* 270:270).

As used herein, the expression "cre gene" means a sequence of nucleotides that encodes a gene product that effects site-specific recombination of DNA in eukaryotic cells at lox sites. One cre gene can be isolated from bacteriophage P1 (see, *e.g.*, Abremski *et al.* (1983) *Cell* 25 32:1301-1311). *E. coli* DH1 and yeast strain BSY90 transformed with plasmid pBS39 carrying a cre gene isolated from bacteriophage P1 and a GAL1 regulatory nucleotide sequence are available from the American Type Culture Collection (ATCC) under accession numbers ATCC 53255

-13-

and ATCC 20772, respectively. The cre gene can be isolated from plasmid pBS39 with restriction enzymes *XhoI* and *SaII*.

As used herein, site-specific recombination refers to site-specific recombination that is effected between two specific sites on a single  
5 nucleic acid molecule or between two different molecules that requires the presence of an exogenous protein, such as an integrase or recombinase.

For example, Cre-lox site-specific recombination can include the following three events:

- 10 a. deletion of a pre-selected DNA segment flanked by lox sites;
- b. inversion of the nucleotide sequence of a pre-selected DNA segment flanked by lox sites; and
- c. reciprocal exchange of DNA segments proximate to  
15 lox sites located on different DNA molecules.

This reciprocal exchange of DNA segments can result in an integration event if one or both of the DNA molecules are circular. DNA segment refers to a linear fragment of single- or double-stranded deoxyribonucleic acid (DNA), which can be derived from any source.

- 20 Since the lox site is an asymmetrical nucleotide sequence, two lox sites on the same DNA molecule can have the same or opposite orientations with respect to each other. Recombination between lox sites in the same orientation results in a deletion of the DNA segment located between the two lox sites and a connection between the resulting ends of the original  
25 DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule result in an inversion of the nucleotide sequence of the DNA segment located between the two lox

-14-

sites. In addition, reciprocal exchange of DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the gene product of the cre gene. Thus, the Cre-lox system can be used to specifically delete, invert, or  
5 insert DNA. The precise event is controlled by the orientation of lox DNA sequences, in *cis* the lox sequences direct the *Cre* recombinase to either delete (lox sequences in direct orientation) or invert (lox sequences in inverted orientation) DNA flanked by the sequences, while *in trans* the lox sequences can direct a homologous recombination event resulting in the  
10 insertion of a recombinant DNA.

As used herein, a chromosome is a nucleic acid molecule, and associated proteins, that is capable of replication and segregation within a cell upon cell division. Typically, a chromosome contains a centromeric region, replication origins, telomeric regions and a region of nucleic acid  
15 between the centromeric and telomeric regions.

As used herein, a centromere is any nucleic acid sequence that confers an ability to segregate to daughter cells through cell division. A centromere may confer stable segregation of a nucleic acid sequence, including an artificial chromosome containing the centromere, through  
20 mitotic or meiotic divisions, including through both mitotic and meiotic divisions. A particular centromere is not necessarily derived from the same species in which it is introduced, but has the ability to promote DNA segregation in cells of that species.

As used herein, euchromatin and heterochromatin have their  
25 recognized meanings. Euchromatin refers to chromatin that stains diffusely and that typically contains genes, and heterochromatin refers to chromatin that remains unusually condensed and that has been thought to be transcriptionally inactive. Highly repetitive DNA sequences (satellite DNA) are usually located in regions of the heterochromatin surrounding

-15-

the centromere (pericentric or pericentromeric heterochromatin).

Constitutive heterochromatin refers to heterochromatin that contains the highly repetitive DNA which is constitutively condensed and genetically inactive.

5       As used herein, an acrocentric chromosome refers to a chromosome with arms of unequal length.

As used herein, endogenous chromosomes refer to genomic chromosomes as found in a cell prior to generation or introduction of an artificial chromosome.

10       As used herein, artificial chromosomes are nucleic acid molecules, typically DNA, that stably replicate and segregate alongside endogenous chromosomes in cells and have the capacity to accommodate and express heterologous genes contained therein. It has the capacity to act as a gene delivery vehicle by accommodating and expressing foreign genes  
15 contained therein. A mammalian artificial chromosome (MAC) refers to chromosomes that have an active mammalian centromere(s). Plant artificial chromosomes, insect artificial chromosomes and avian artificial chromosomes refer to chromosomes that include centromeres that function in plant, insect and avian cells, respectively. A human artificial  
20 chromosome (HAC) refers to chromosomes that include centromeres that function in human cells. For exemplary artificial chromosomes, see, *e.g.*, U.S. Patent Nos. 6,025,155; 6,077,697; 5,288,625; 5,712,134; 5,695,967; 5,869,294; 5,891,691 and 5,721,118 and published International PCT application Nos, WO 97/40183 and WO 98/08964.  
25 Artificial chromosomes include those that are predominantly heterochromatic (formerly referred to as satellite artificial chromosomes (SATACs); see, *e.g.*, U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183), minichromosomes that contain a *de novo* centromere (see, U.S. Patent

-16-

Nos. 5,712,134, 5,891,691 and 5,288,625), artificial chromosomes predominantly made up of repeating nucleic acid units and that contain substantially equivalent amounts of euchromatic and heterochromatic DNA and *in vitro* assembled artificial chromosomes (see, copending U.S. provisional application Serial No. 60/294,687, filed on May 30, 2001).

As used herein, the term "satellite DNA-based artificial chromosome (SATAC)" is interchangeable with the term "artificial chromosome expression system (ACes)". These artificial chromosomes (ACes) include those that are substantially all neutral non-coding sequences (heterochromatin) except for foreign heterologous, typically gene-encoding nucleic acid, that is interspersed within the heterochromatin for the expression therein (see U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183), or that is in a single locus as provided herein. Also included are ACes that may include euchromatin and that result from the process described in U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183 and outlined herein. The delineating structural feature is the presence of repeating units, that are generally predominantly heterochromatin. The precise structure of the ACes will depend upon the structure of the chromosome in which the initial amplification event occurs; all share the common feature of including a defined pattern of repeating units. Generally ACes have more heterochromatin than euchromatin. Foreign nucleic acid molecules (heterologous genes) contained in these artificial chromosome expression systems can include any nucleic acid whose expression is of interest in a particular host cell. Such foreign nucleic acid molecules, include, but are not limited to, nucleic acid that encodes traceable marker proteins (reporter genes), such as fluorescent proteins, such as green, blue or red fluorescent proteins (GFP, BFP and RFP, respectively), other reporter

-17-

genes, such as  $\beta$ -galactosidase and proteins that confer drug resistance, such as a gene encoding hygromycin-resistance. Other examples of heterologous nucleic acid molecules include, but are not limited to, DNA that encodes therapeutically effective substances, such as anti-cancer  
5 agents, enzymes and hormones, DNA that encodes other types of proteins, such as antibodies, and DNA that encodes RNA molecules (such as antisense or siRNA molecules) that are not translated into proteins.

As used herein, an artificial chromosome platform, also referred to  
10 herein as a "platform ACes" or "ACes platform", refers to an artificial chromosome that has been engineered to include one or more sites for site-specific, recombination-directed integration. In particular, ACes that are so-engineered are provided. Any sites, including but not limited to any described herein, that are suitable for such integration are  
15 contemplated. Plant and animal platform ACes are provided. Among the ACes contemplated herein are those that are predominantly heterochromatic (formerly referred to as satellite artificial chromosomes (SATACs); see, *e.g.*, U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183), artificial  
20 chromosomes predominantly made up of repeating nucleic acid units and that contain substantially equivalent amounts of euchromatic and heterochromatic DNA resulting from an amplification event depicted in the referenced patent and herein. Included among the ACes for use in generating platforms, are artificial chromosomes that introduce and  
25 express heterologous nucleic acids in plants (see, copending U.S. provisional application Serial No. 60/294,687, filed on May 30, 2001). These include artificial chromosomes that have a centromere derived from a plant, and, also, artificial chromosomes that have centromeres that may be derived from other organisms but that function in plants.



-18-

As used herein a "reporter ACes" refers to a an ACes that comprises one or a plurality of reporter constructs, where the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

5 As used herein, amplification, with reference to DNA, is a process in which segments of DNA are duplicated to yield two or multiple copies of substantially similar or identical or nearly identical DNA segments that are typically joined as substantially tandem or successive repeats or inverted repeats.

10 As used herein, amplification-based artificial chromosomes are artificial chromosomes derived from natural or endogenous chromosomes by virtue of an amplification event, such as one initiated by introduction of heterologous nucleic acid into rDNA in a chromosome. As a result of such an event, chromosomes and fragments thereof exhibiting segmented  
15 or repeating patterns arise. Artificial chromosomes can be formed from these chromosomes and fragments. Hence, amplification-based artificial chromosomes refer to engineered chromosomes that exhibit an ordered segmentation that is not observed in naturally occurring chromosomes and that distinguishes them from naturally occurring chromosomes. The  
20 segmentation, which can be visualized using a variety of chromosome analysis techniques known to those of skill in the art, correlates with the structure of these artificial chromosomes. In addition to containing one or more centromeres, the amplification-based artificial chromosomes, throughout the region or regions of segmentation are predominantly made  
25 up of nucleic acid units also referred to as "amplicons", that is (are) repeated in the region and that have a similar gross structure. Repeats of an amplicon tend to be of similar size and share some common nucleic acid sequences. For example, each repeat of an amplicon may contain a replication site involved in amplification of chromosome segments and/or

-19-

some heterologous nucleic acid that was utilized in the initial production of the artificial chromosome. Typically, the repeating units are substantially similar in nucleic acid composition and may be nearly identical.

5           The amplification-based artificial chromosomes differ depending on the chromosomal region that has undergone amplification in the process of artificial chromosome formation. The structures of the resulting chromosomes can vary depending upon the initiating event and/or the conditions under which the heterologous nucleic acid is introduced,  
10 including modification to the endogenous chromosomes. For example, in some of the artificial chromosomes provided herein, the region or regions of segmentation may be made up predominantly of heterochromatic DNA. In other artificial chromosomes provided herein, the region or regions of segmentation may be made up predominantly of euchromatic DNA or may  
15 be made up of similar amounts of heterochromatic and euchromatic DNA.

As used herein an amplicon is a repeated nucleic acid unit. In some of the artificial chromosomes described herein, an amplicon may contain a set of inverted repeats of a megareplicon. A megareplicon represents a higher order replication unit. For example, with reference to  
20 some of the predominantly heterochromatic artificial chromosomes, the megareplicon can contain a set of tandem DNA blocks (*e.g.*, ~7.5 Mb DNA blocks) each containing satellite DNA flanked by non-satellite DNA or may be made up of substantially rDNA. Contained within the megareplicon is a primary replication site, referred to as the  
25 megareplicator, which may be involved in organizing and facilitating replication of the pericentric heterochromatin and possibly the centromeres. Within the megareplicon there may be smaller (*e.g.*, 50-300 kb) secondary replicons.

-20-

In artificial chromosomes, such as those provided U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183, the megareplicon is defined by two tandem blocks (~7.5 Mb DNA blocks in the chromosomes provided therein). Within each artificial  
5 chromosome or among a population thereof, each amplicon has the same gross structure but may contain sequence variations. Such variations will arise as a result of movement of mobile genetic elements, deletions or insertions or mutations that arise, particularly in culture. Such variation does not affect the use of the artificial chromosomes or their overall  
10 structure as described herein.

As used herein, amplifiable, when used in reference to a chromosome, particularly the method of generating artificial chromosomes provided herein, refers to a region of a chromosome that is prone to amplification. Amplification typically occurs during replication and other  
15 cellular events involving recombination (*e.g.*, DNA repair). Such regions include regions of the chromosome that contain tandem repeats, such as satellite DNA, rDNA, and other such sequences.

As used herein, a dicentric chromosome is a chromosome that contains two centromeres. A multicentric chromosome contains more  
20 than two centromeres.

As used herein, a formerly dicentric chromosome is a chromosome that is produced when a dicentric chromosome fragments and acquires new telomeres so that two chromosomes, each having one of the centromeres, are produced. Each of the fragments is a replicable  
25 chromosome. If one of the chromosomes undergoes amplification of primarily euchromatic DNA to produce a fully functional chromosome that is predominantly (at least more than 50%) euchromatin, it is a minichromosome. The remaining chromosome is a formerly dicentric chromosome. If one of the chromosomes undergoes amplification,

-21-

whereby heterochromatin (such as, for example, satellite DNA) is amplified and a euchromatic portion (such as, for example, an arm) remains, it is referred to as a sausage chromosome. A chromosome that is substantially all heterochromatin, except for portions of heterologous DNA, is called a predominantly heterochromatic artificial chromosome. Predominantly heterochromatic artificial chromosomes can be produced from other partially heterochromatic artificial chromosomes by culturing the cell containing such chromosomes under conditions such as BrdU treatment that destabilize the chromosome and/or growth under selective conditions so that a predominantly heterochromatic artificial chromosome is produced. For purposes herein, it is understood that the artificial chromosomes may not necessarily be produced in multiple steps, but may appear after the initial introduction of the heterologous DNA. Typically, artificial chromosomes appear after about 5 to about 60, or about 5 to about 55, or about 10 to about 55 or about 25 to about 55 or about 35 to about 55 cell doublings after initiation of artificial chromosome generation, or they may appear after several cycles of growth under selective conditions and BrdU treatment.

As used herein, an artificial chromosome that is predominantly heterochromatic (*i.e.*, containing more heterochromatin than euchromatin, typically more than about 50%, more than about 70%, or more than about 90% heterochromatin) may be produced by introducing nucleic acid molecules into cells, such as, for example, animal or plant cells, and selecting cells that contain a predominantly heterochromatic artificial chromosome. Any nucleic acid may be introduced into cells in such methods of producing the artificial chromosomes. For example, the nucleic acid may contain a selectable marker and/or optionally a sequence that targets nucleic acid to the pericentric, heterochromatic region of a chromosome, such as in the short arm of acrocentric chromosomes and

-22-

nucleolar organizing regions. Targeting sequences include, but are not limited to, lambda phage DNA and rDNA for production of predominantly heterochromatic artificial chromosomes in eukaryotic cells.

After introducing the nucleic acid into cells, a cell containing a  
5 predominantly heterochromatic artificial chromosome is selected. Such cells may be identified using a variety of procedures. For example, repeating units of heterochromatic DNA of these chromosomes may be discerned by G-banding and/or fluorescence in situ hybridization (FISH) techniques. Prior to such analyses, the cells to be analyzed may be  
10 enriched with artificial chromosome-containing cells by sorting the cells on the basis of the presence of a selectable marker, such as a reporter protein, or by growing (culturing) the cells under selective conditions. It is also possible, after introduction of nucleic acids into cells, to select cells that have a multicentric, typically dicentric, chromosome, a formerly  
15 multicentric (typically dicentric) chromosome and/or various heterochromatic structures, such as a megachromosome and a sausage chromosome, that contain a centromere and are predominantly heterochromatic and to treat them such that desired artificial chromosomes are produced. Cells containing a new chromosome are  
20 selected. Conditions for generation of a desired structure include, but are not limited to, further growth under selective conditions, introduction of additional nucleic acid molecules and/or growth under selective conditions and treatment with destabilizing agents, and other such methods (see International PCT application No. WO 97/40183 and U.S. Patent Nos.  
25 6,025,155 and 6,077,697).

As used herein, a "selectable marker" is a nucleic acid segment, generally DNA, that allows one to select for or against a molecule or a cell that contains it, often under particular conditions. These markers can encode an activity, such as, but not limited to, production of RNA,

-23-

peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds and compositions. Examples of selectable markers include but are not limited to: (1) nucleic acid segments that encode products that provide resistance against otherwise  
5 toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid segments that encode products that can be identified, such as phenotypic markers,  
10 including  $\beta$ -galactosidase, red, blue and/or green fluorescent proteins (FPs), and cell surface proteins; (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described in Nos. 1-5 above (e.g., antisense  
15 oligonucleotides or siRNA molecules for use in RNA interference); (7) nucleic acid segments that bind products that modify a substrate (e.g. restriction endonucleases); (8) nucleic acid segments that can be used to isolate a desired molecule (e.g. specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence that can be  
20 otherwise non-functional, such as for PCR amplification of subpopulations of molecules; and/or (10) nucleic acid segments, which when absent, directly or indirectly confer sensitivity to particular compounds. Thus, for example, selectable markers include nucleic acids encoding fluorescent proteins, such as green fluorescent proteins,  $\beta$ -galactosidase and other  
25 readily detectable proteins, such as chromogenic proteins or proteins capable of being bound by an antibody and FACs sorted. Selectable markers such as these, which are not required for cell survival and/or proliferation in the presence of a selection agent, are also referred to herein as reporter molecules. Other selectable markers, *e.g.*, the

-24-

neomycin phosphotransferase gene, provide for isolation and identification of cells containing them by conferring properties on the cells that make them resistant to an agent, *e.g.*, a drug such as an antibiotic, that inhibits proliferation of cells that do not contain the marker.

- 5           As another example, interference of gene expression by double stranded RNA has been shown in *Caenorhabditis elegans*, plants, *Drosophila*, protozoans and mammals. This method is known as RNA interference (RNAi) and utilizes short, double-stranded RNA molecules (siRNAs). The siRNAs are generally composed of a 19-22bp double-
- 10   stranded RNA stem, a loop region and a 1-4 bp overhang on the 3' end. The reduction of gene expression has been accomplished by direct introduction of the siRNAs into the cell (Harborth J et al., 2001, *J Cell Sci* 114(pt 24):4557-65) as well as the introduction of DNA encoding and expressing the siRNA molecule. The encoded siRNA molecules are under
- 15   the regulation of an RNA polymerase III promoter (see, *e.g.*, Yu et al., 2002, *Proc Natl Acad Sci USA* 99(9):6047-52; Brummelkamp et al., 2002, *Science* 296(5567):550-3; Miyagishi et al., 2002, *Nat Biotechnol* 20(5):497-500; and the like). In certain embodiments, RNAi in mammalian cells may have advantages over other therapeutic methods.
- 20   For example, producing siRNA molecules that block viral genetic activities in infected cells may reduce the effects of the virus. Platform ACes provided herein encoding siRNA molecule(s) are an additional utilization of the platform ACes technology. The platform ACes could be engineered to encode one or more siRNA molecules to create gene "knockdowns". In
- 25   one embodiment, a platform ACes can engineered to encode both the siRNA molecule and a replacement gene. For example, a mouse model or cell culture system could be generated using a platform ACes that has a knockdown of the endogenous mouse gene, by siRNA, and the human gene homolog expressing in place of the mouse gene. The placement of

-25-

siRNA encoding sequences under the regulation of a regulatable or inducible promoter would allow one to temporally and/or spatially control the knockdown effect of the corresponding gene.

As used herein, a reporter gene includes any gene that expresses a  
5 detectable gene product, which may be RNA or protein. Generally  
reporter genes are readily detectable. Examples of reporter genes include,  
but are not limited to nucleic acid encoding a fluorescent protein, CAT  
(chloramphenicol acetyl transferase) (Alton *et al.* (1979) *Nature* 282: 864-  
869) luciferase, and other enzyme detection systems, such as beta-  
10 galactosidase; firefly luciferase (deWet *et al.* (1987) *Mol. Cell. Biol.*  
7:725-737); bacterial luciferase (Engebrecht and Silverman (1984) *Proc.*  
*Natl. Acad. Sci. U.S.A.* 81:4154-4158; Baldwin *et al.* (1984)  
*Biochemistry* 23:3663-3667); and alkaline phosphatase (Toh *et al.* (1989)  
*Eur. J. Biochem.* 182:231-238, Hall *et al.* (1983) *J. Mol. Appl. Gen.*  
15 2:101).

As used herein, growth under selective conditions means growth of  
a cell under conditions that require expression of a selectable marker for  
survival.

As used herein, an agent that destabilizes a chromosome is any  
20 agent known by those skilled in the art to enhance amplification events,  
and/or mutations. Such agents, which include BrdU, are well known to  
those skilled in the art.

In order to generate an artificial chromosome containing a particular  
heterologous nucleic acid of interest, it is possible to include the nucleic  
25 acid in the nucleic acid that is being introduced into cells to initiate  
production of the artificial chromosome. Thus, for example, a nucleic  
acid can be introduced into a cell along with nucleic acid encoding a  
selectable marker and/or a nucleic acid that targets to a heterochromatic  
region of a chromosome. For introducing a heterologous nucleic acid into



-26-

the cell, it can be included in a fragment that includes a selectable marker or as part of a separate nucleic acid fragment and introduced into the cell with a selectable marker during the process of generating the artificial chromosomes. Alternatively, heterologous nucleic acid can be introduced  
5 into an artificial chromosome at a later time after the initial generation of the artificial chromosome.

As used herein, the minichromosome refers to a chromosome derived from a multicentric, typically dicentric, chromosome that contains more euchromatic than heterochromatic DNA. For purposes herein, the  
10 minichromosome contains a *de novo* centromere (e.g., a neocentromere). In some embodiments, for example, the minichromosome contains a centromere that replicates in animals, e.g., a mammalian centromere or in plants, e.g., a plant centromere.

As used herein, *in vitro* assembled artificial chromosomes or  
15 synthetic chromosomes can be either more euchromatic than heterochromatic or more heterochromatic than euchromatic and are produced by joining essential components of a chromosome *in vitro*. These components include at least a centromere, a megareplicator, a telomere and optionally secondary origins of replication.

20 As used herein, *in vitro* assembled plant or animal artificial chromosomes are produced by joining essential components (at least the centromere, telomere(s), megareplicator and optional secondary origins of replication) that function in plants or animals. In particular embodiments, the megareplicator contains sequences of rDNA, particularly plant or  
25 animal rDNA.

As used herein, a plant is a eukaryotic organism that contains, in addition to a nucleus and mitochondria, chloroplasts capable of carrying out photosynthesis. A plant can be unicellular or multicellular and can contain multiple tissues and/or organs. Plants can reproduce sexually or

-27-

asexually and can be perennial or annual in growth. Plants can also be terrestrial or aquatic. The term "plant" includes a whole plant, plant cell, plant protoplast, plant calli, plant seed, plant organ, plant tissue, and other parts of a whole plant.

- 5           As used herein, stable maintenance of chromosomes occurs when at least about 85%, preferably 90%, more preferably 95%, of the cells retain the chromosome. Stability is measured in the presence of a selective agent. Preferably these chromosomes are also maintained in the absence of a selective agent. Stable chromosomes also retain their  
10 structure during cell culturing, suffering no unintended intrachromosomal or interchromosomal rearrangements.

- As used herein, *de novo* with reference to a centromere, refers to generation of an excess centromere in a chromosome as a result of incorporation of a heterologous nucleic acid fragment using the methods  
15 herein.

- As used herein, BrdU refers to 5-bromodeoxyuridine, which during replication is inserted in place of thymidine. BrdU is used as a mutagen; it also inhibits condensation of metaphase chromosomes during cell division.

- 20           As used herein, ribosomal RNA (rRNA) is the specialized RNA that forms part of the structure of a ribosome and participates in the synthesis of proteins. Ribosomal RNA is produced by transcription of genes which, in eukaryotic cells, are present in multiple copies. In human cells, the approximately 250 copies of rRNA genes (i.e., genes which encode rRNA)  
25 per haploid genome are spread out in clusters on at least five different chromosomes (chromosomes 13, 14, 15, 21 and 22). In mouse cells, the presence of ribosomal DNA (rDNA, which is DNA containing sequences that encode rRNA) has been verified on at least 11 pairs out of 20 mouse chromosomes (chromosomes 5, 6, 7, 9, 11, 12, 15, 16, 17, 18, and 19)

-28-

(see e.g., Rowe *et al.* (1996) *Mamm. Genome* 7:886-889 and Johnson *et al.* (1993) *Mamm. Genome* 4:49-52). In *Arabidopsis thaliana* the presence of rDNA has been verified on chromosomes 2 and 4 (18S, 5.8S, and 25S rDNA) and on chromosomes 3,4, and 5 (5S rDNA)(see The

5 Arabidopsis Genome Initiative (2000) *Nature* 408:796-815). In eukaryotic cells, the multiple copies of the highly conserved rRNA genes are located in a tandemly arranged series of rDNA units, which are generally about 40-45 kb in length and contain a transcribed region and a nontranscribed region known as spacer (i.e., intergenic spacer) DNA

10 which can vary in length and sequence. In the human and mouse, these tandem arrays of rDNA units are located adjacent to the pericentric satellite DNA sequences (heterochromatin). The regions of these chromosomes in which the rDNA is located are referred to as nucleolar organizing regions (NOR) which loop into the nucleolus, the site of

15 ribosome production within the cell nucleus.

As used herein, a megachromosome refers to a chromosome that, except for introduced heterologous DNA, is substantially composed of heterochromatin. Megachromosomes are made up of an array of repeated amplicons that contain two inverted megareplicons bordered by

20 introduced heterologous DNA (see, *e.g.*, Figure 3 of U.S. Patent No. 6,077,697 for a schematic drawing of a megachromosome). For purposes herein, a megachromosome is about 50 to 400 Mb, generally about 250-400 Mb. Shorter variants are also referred to as truncated megachromosomes (about 90 to 120 or 150 Mb), dwarf

25 megachromosomes (~150-200 Mb), and a micro-megachromosome (~50-90 Mb, typically 50-60 Mb). For purposes herein, the term

-29-

megachromosome refers to the overall repeated structure based on an array of repeated chromosomal segments (amplicons) that contain two inverted megareplicons bordered by any inserted heterologous DNA. The size will be specified.

5           As used herein, gene therapy involves the transfer or insertion of nucleic acid molecules into certain cells, which are also referred to as target cells, to produce specific products that are involved in preventing, curing, correcting, controlling or modulating diseases, disorders and deleterious conditions. The nucleic acid is introduced into the selected  
10 target cells in a manner such that the nucleic acid is expressed and a product encoded thereby is produced. Alternatively, the nucleic acid may in some manner mediate expression of DNA that encodes a therapeutic product. This product may be a therapeutic compound, which is produced in therapeutically effective amounts or at a therapeutically  
15 useful time. It may also encode a product, such as a peptide or RNA, that in some manner mediates, directly or indirectly, expression of a therapeutic product. Expression of the nucleic acid by the target cells within an organism afflicted with a disease or disorder thereby provides for modulation of the disease or disorder. The nucleic acid encoding the  
20 therapeutic product may be modified prior to introduction into the cells of the afflicted host in order to enhance or otherwise alter the product or expression thereof.

For use in gene therapy, cells can be transfected *in vitro*, followed by introduction of the transfected cells into an organism. This is often  
25 referred to as *ex vivo* gene therapy. Alternatively, the cells can be transfected directly *in vivo* within an organism.

As used herein, therapeutic agents include, but are not limited to, growth factors, antibodies, cytokines, such as tumor necrosis factors and interleukins, and cytotoxic agents and other agents disclosed herein and

-30-

known to those of skill in the art. Such agents include, but are not limited to, tumor necrosis factor,  $\alpha$ -interferon,  $\beta$ -interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 (IL-1), interleukin-2 (IL-2), interleukin-6 (IL-6), granulocyte macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), erythropoietin (EPO), pro-coagulants such as tissue factor and tissue factor variants, pro-apoptotic agents such as FAS-ligand, fibroblast growth factors (FGF), nerve growth factor and other growth factors.

As used herein, a therapeutically effective product is a product that is encoded by heterologous DNA that, upon introduction of the DNA into a host, a product is expressed that effectively ameliorates or eliminates the symptoms, manifestations of an inherited or acquired disease or that cures the disease.

As used herein, transgenic plants and animals refer to plants and animals in which heterologous or foreign nucleic acid is expressed or in which the expression of a gene naturally present in the plant or animal has been altered by virtue of introduction of heterologous or foreign nucleic acid.

As used herein, IRES (internal ribosome entry site; see, *e.g.*, SEQ ID No. 27 and nucleotides 2736-3308 SEQ ID No. 28) refers to a region of a nucleic acid molecule, such as an mRNA molecule, that allows internal ribosome entry sufficient to initiate translation, which initiation can be detected in an assay for cap-independent translation (see, *e.g.*, U.S. Patent No. 6,171,821). The presence of an IRES within an mRNA molecule allows cap-independent translation of a linked protein-encoding sequence that otherwise would not be translated.

-31-

Internal ribosome entry site (IRES) elements were first identified in picornaviruses, which elements are considered the paradigm for cap-independent translation. The 5' UTRs of all picornaviruses are long and mediate translational initiation by directly recruiting and binding  
5 ribosomes, thereby circumventing the initial cap-binding step. IRES elements are frequently found in viral mRNA, they are rare in non-viral mRNA. Among non-viral mRNA molecules that contain functional IRES elements in their respective 5' UTRs are those encoding immunoglobulin heavy chain binding protein (BiP) (Macejak *et al.* (1991) *Nature*  
10 353:90-94); *Drosophila* Antennapedia (Oh *et al.* (1992) *Genes Dev.* 6:1643-1653); *D. Ultrabithorax* (Ye *et al.* (1997) *Mol. Cell Biol.* 17:1714-21); fibroblast growth factor 2 (Vagner *et al.* (1995) *Mol. Cell Biol.* 15:35-44); initiation factor eIF4G (Gan *et al.* (1998) *J. Biol. Chem.* 273:5006-5012); proto-oncogene c-myc (Nanbru *et al.* (1995) *J. Biol.*  
15 *Chem.* 272:32061-32066; Stoneley (1998) *Oncogene* 16:423-428); IRES<sub>H</sub> from the 5'UTR of NRF1 gene (Oumard *et al.* (2000) *Mol. and Cell Biol.*, 20(8):2755-2759); and vascular endothelial growth factor (VEGF) (Stein *et al.* (1998) *Mol. Cell Biol.* 18:3112-9).

As used herein, a promoter, with respect to a region of DNA, refers  
20 to a sequence of DNA that contains a sequence of bases that signals RNA polymerase to associate with the DNA and initiate transcription of RNA (such as pol II for mRNA) from a template strand of the DNA. A promoter thus generally regulates transcription of DNA into mRNA. A particular promoter provided herein is the Ferritin heavy chain promoter (excluding  
25 the Iron Response Element, located in the 5'UTR), which was joined to the 37bp Fer-1 enhancer element. This promoter is set forth as SEQ ID NO:128. The endogenous Fer-1 enhancer element is located upstream of the Fer-1 promoter (e.g., a Fer-1 oligo was cloned proximal to the core promoter).

-32-

As used herein, isolated, substantially pure nucleic acid, such as, for example, DNA, refers to nucleic acid fragments purified according to standard techniques employed by those skilled in the art, such as that found in Sambrook *et al.* ((2001) Molecular Cloning: A Laboratory  
5 Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 3rd edition).

As used herein, expression refers to the transcription and/or translation of nucleic acid. For example, expression can be the transcription of a gene that may be transcribed into an RNA molecule,  
10 such as a messenger RNA (mRNA) molecule. Expression may further include translation of an RNA molecule and translated into peptides, polypeptides, or proteins. If the nucleic acid is derived from genomic DNA, expression may, if an appropriate eukaryotic host cell or organism is selected, include splicing of the mRNA. With respect to an antisense  
15 construct, expression may refer to the transcription of the antisense DNA.

As used herein, vector or plasmid refers to discrete elements that are used to introduce heterologous nucleic acids into cells for either expression of the heterologous nucleic acid or for replication of the heterologous nucleic acid. Selection and use of such vectors and  
20 plasmids are well within the level of skill of the art.

As used herein, transformation/transfection refers to the process by which nucleic acid is introduced into cells. The terms transfection and transformation refer to the taking up of exogenous nucleic acid, e.g., an expression vector, by a host cell whether or not any coding sequences  
25 are in fact expressed. Numerous methods of transfection are known to the ordinarily skilled artisan, for example, by *Agrobacterium*-mediated transformation, protoplast transformation (including polyethylene glycol (PEG)-mediated transformation, electroporation, protoplast fusion, and microcell fusion), lipid-mediated delivery, liposomes, electroporation,

-33-

sonoporation, microinjection, particle bombardment and silicon carbide whisker-mediated transformation and combinations thereof (see, *e.g.*, Paszkowski *et al.* (1984) *EMBO J.* 3:2717-2722; Potrykus *et al.* (1985) *Mol. Gen. Genet.* 199:169-177; Reich *et al.* (1986) *Biotechnology* 4:1001-1004; Klein *et al.* (1987) *Nature* 327:70-73; U.S. Patent No. 6,143,949; Paszkowski *et al.* (1989) in *Cell Culture and Somatic Cell Genetics of Plants, Vol. 6*, Molecular Biology of Plant Nuclear Genes, eds. Schell, J and Vasil, L.K. Academic Publishers, San Diego, California, p. 52-68; and Frame *et al.* (1994) *Plant J.* 6:941-948), direct uptake using calcium phosphate (CaPO<sub>4</sub>; see, *e.g.*, Wigler *et al.* (1979) *Proc. Natl. Acad. Sci. U.S.A.* 76:1373-1376), polyethylene glycol (PEG)-mediated DNA uptake, lipofection (see, *e.g.*, Strauss (1996) *Meth. Mol. Biol.* 54:307-327), microcell fusion (see, EXAMPLES, see, also Lambert (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:5907-5911; U.S. Patent No. 5,396,767, Sawford *et al.* (1987) *Somatic Cell Mol. Genet.* 13:279-284; Dhar *et al.* (1984) *Somatic Cell Mol. Genet.* 10:547-559; and McNeill-Killary *et al.* (1995) *Meth. Enzymol.* 254:133-152), lipid-mediated carrier systems (see, *e.g.*, Teifel *et al.* (1995) *Biotechniques* 19:79-80; Albrecht *et al.* (1996) *Ann. Hematol.* 72:73-79; Holmen *et al.* (1995) *In Vitro Cell Dev. Biol. Anim.* 31:347-351; Remy *et al.* (1994) *Bioconjug. Chem.* 5:647-654; Le Bolch *et al.* (1995) *Tetrahedron Lett.* 36:6681-6684; Loeffler *et al.* (1993) *Meth. Enzymol.* 217:599-618) or other suitable method. Methods for delivery of ACes are described in copending U.S. application Serial No. 09/815,979. Successful transfection is generally recognized by detection of the presence of the heterologous nucleic acid within the transfected cell, such as, for example, any visualization of the heterologous nucleic acid or any indication of the operation of a vector within the host cell.



-34-

As used herein, "delivery," which is used interchangeably with "transfection," refers to the process by which exogenous nucleic acid molecules are transferred into a cell such that they are located inside the cell. Delivery of nucleic acids is a distinct process from expression of  
5 nucleic acids.

As used herein, injected refers to the microinjection, such as by use of a small syringe, needle, or pipette, for injection of nucleic acid into a cell.

As used herein, substantially homologous DNA refers to DNA that  
10 includes a sequence of nucleotides that is sufficiently similar to another such sequence to form stable hybrids, with each other or a reference sequence, under specified conditions.

It is well known to those of skill in this art that nucleic acid fragments with different sequences may, under the same conditions,  
15 hybridize detectably to the same "target" nucleic acid. Two nucleic acid fragments hybridize detectably, under stringent conditions over a sufficiently long hybridization period, because one fragment contains a segment of at least about 10, 14 or 16 or more nucleotides in a sequence that is complementary (or nearly complementary) to a substantially  
20 contiguous sequence of at least one segment in the other nucleic acid fragment. If the time during which hybridization is allowed to occur is held constant, at a value during which, under preselected stringency conditions, two nucleic acid fragments with complementary base-pairing segments hybridize detectably to each other, departures from exact  
25 complementarity can be introduced into the base-pairing segments, and base-pairing will nonetheless occur to an extent sufficient to make hybridization detectable. As the departure from complementarity between the base-pairing segments of two nucleic acids becomes larger, and as

-35-

conditions of the hybridization become more stringent, the probability decreases that the two segments will hybridize detectably to each other.

Two single-stranded nucleic acid segments have "substantially the same sequence", if (a) both form a base-paired duplex with the same  
5 segment, and (b) the melting temperatures of the two duplexes in a solution of 0.5 X SSPE differ by less than 10°C. If the segments being compared have the same number of bases, then to have "substantially the same sequence", they will typically differ in their sequences at fewer than 1 base in 10. Methods for determining melting temperatures of  
10 nucleic acid duplexes are well known (see, *e.g.*, Meinkoth *et al.* (1984) *Anal. Biochem.* 138:267-284 and references cited therein).

As used herein, a nucleic acid probe is a DNA or RNA fragment that includes a sufficient number of nucleotides to specifically hybridize to DNA or RNA that includes complementary or substantially complementary  
15 sequences of nucleotides. A probe may contain any number of nucleotides, from as few as about 10 and as many as hundreds of thousands of nucleotides. The conditions and protocols for such hybridization reactions are well known to those of skill in the art as are the effects of probe size, temperature, degree of mismatch, salt  
20 concentration and other parameters on the hybridization reaction. For example, the lower the temperature and higher the salt concentration at which the hybridization reaction is carried out, the greater the degree of mismatch that may be present in the hybrid molecules.

To be used as a hybridization probe, the nucleic acid is generally  
25 rendered detectable by labeling it with a detectable moiety or label, such as  $^{32}\text{P}$ ,  $^3\text{H}$  and  $^{14}\text{C}$ , or by other means, including chemical labeling, such as by nick-translation in the presence of deoxyuridylate biotinylated at the 5'-position of the uracil moiety. The resulting probe includes the biotinylated uridylate in place of thymidylate residues and can be detected

-36-

(via the biotin moieties) by any of a number of commercially available detection systems based on binding of streptavidin to the biotin. Such commercially available detection systems can be obtained, for example, from Enzo Biochemicals, Inc. (New York, NY). Any other label known to  
5 those of skill in the art, including non-radioactive labels, may be used as long as it renders the probes sufficiently detectable, which is a function of the sensitivity of the assay, the time available (for culturing cells, extracting DNA, and hybridization assays), the quantity of DNA or RNA available as a source of the probe, the particular label and the means used  
10 to detect the label.

Once sequences with a sufficiently high degree of homology to the probe are identified, they can readily be isolated by standard techniques (see, *e.g.*, Sambrook *et al.* (2001) *Molecular Cloning: A Laboratory Manual*, 3rd Edition, Cold Spring Harbor Laboratory Press).

15 As used herein, conditions under which DNA molecules form stable hybrids are considered substantially homologous, and a DNA or nucleic acid homolog refers to a nucleic acid that includes a preselected conserved nucleotide sequence, such as a sequence encoding a polypeptide. By the term "substantially homologous" is meant having at  
20 least 75%, preferably 80%, preferably at least 90%, most preferably at least 95% homology therewith or a less percentage of homology or identity and conserved biological activity or function.

The terms "homology" and "identity" are often used interchangeably. In this regard, percent homology or identity may be  
25 determined, for example, by comparing sequence information using a GAP computer program. The GAP program utilizes the alignment method of Needleman and Wunsch (*J. Mol. Biol.* 48:443 (1970)), as revised by Smith and Waterman (*Adv. Appl. Math.* 2:482 (1981)). Briefly, the GAP program defines similarity as the number of aligned symbols (*i.e.*,

-37-

nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences. The preferred default parameters for the GAP program may include: (1) a unary comparison matrix (containing a value of 1 for identities and 0 for non-identities) and  
5 the weighted comparison matrix of Gribskov and Burgess, *Nucl. Acids Res.* 14:6745 (1986), as described by Schwartz and Dayhoff, eds., *ATLAS OF PROTEIN SEQUENCE AND STRUCTURE*, National Biomedical Research Foundation, pp. 353-358 (1979); (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3)  
10 no penalty for end gaps.

By sequence identity, the number of conserved amino acids are determined by standard alignment algorithms programs, and are used with default gap penalties established by each supplier. Substantially homologous nucleic acid molecules would hybridize typically at moderate  
15 stringency or at high stringency all along the length of the nucleic acid of interest. Preferably the two molecules will hybridize under conditions of high stringency. Also contemplated are nucleic acid molecules that contain degenerate codons in place of codons in the hybridizing nucleic acid molecule.

20 Whether any two nucleic acid molecules have nucleotide sequences that are at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% "identical" can be determined using known computer algorithms such as the "FAST A" program, using for example, the default parameters as in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 (1988).  
25 Alternatively the BLAST function of the National Center for Biotechnology Information database may be used to determine relative sequence identity.

In general, sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be

-38-

calculated using published techniques. (See, *e.g.*: *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H. & Lipton, D., *SIAM J Applied Math* 48:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H. & Lipton, D., *SIAM J Applied Math* 48:1073 (1988). Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., *et al.*, *Nucleic Acids Research* 12(1):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., *et al.*, *J Molec Biol* 215:403 (1990)).

Therefore, as used herein, the term "identity" represents a comparison between a test and a reference polypeptide or polynucleotide. For example, a test polypeptide may be defined as any polypeptide that is 90% or more identical to a reference polypeptide.

As used herein, the term at least "90% identical to" refers to percent identities from 90 to 99.99 relative to the reference polypeptides. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of

-39-

100 amino acids are compared. No more than 10% (i.e., 10 out of 100) amino acids in the test polypeptide differs from that of the reference polypeptides. Similar comparisons may be made between a test and reference polynucleotides. Such differences may be represented as point mutations randomly distributed over the entire length of an amino acid sequence or they may be clustered in one or more locations of varying length up to the maximum allowable, e.g. 10/100 amino acid difference (approximately 90% identity). Differences are defined as nucleic acid or amino acid substitutions, or deletions.

5

10 As used herein: stringency of hybridization in determining percentage mismatch encompass the following conditions or equivalent conditions thereto:

- 1) high stringency: 0.1 x SSPE or SSC, 0.1% SDS, 65°C
- 2) medium stringency: 0.2 x SSPE or SSC, 0.1% SDS, 50°C
- 15 3) low stringency: 1.0 x SSPE or SSC, 0.1% SDS, 50°C

or any combination of salt and temperature and other reagents that result in selection of the same degree of mismatch or matching. Equivalent conditions refer to conditions that select for substantially the same percentage of mismatch in the resulting hybrids. Additions of ingredients, such as formamide, Ficoll, and Denhardt's solution affect parameters such as the temperature under which the hybridization should be conducted and the rate of the reaction. Thus, hybridization in 5 X SSC, in 20% formamide at 42° C is substantially the same as the conditions recited above hybridization under conditions of low stringency. The recipes for 25 SSPE, SSC and Denhardt's and the preparation of deionized formamide are described, for example, in Sambrook *et al.* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Chapter 8; see, Sambrook *et al.*, vol. 3, p. B.13, see, also, numerous catalogs that describe commonly used laboratory solutions. It is understood that

-40-

equivalent stringencies may be achieved using alternative buffers, salts and temperatures. As used herein, all assays and procedures, such as hybridization reactions and antibody-antigen reactions, unless otherwise specified, are conducted under conditions recognized by those of skill in the art as standard conditions.

As used herein, conservative amino acid substitutions, such as those set forth in Table 1, are those that do not eliminate biological activity. Suitable conservative substitutions of amino acids are known to those of skill in this art and may be made generally without altering the biological activity of the resulting molecule. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (see, *e.g.*, Watson *et al. Molecular Biology of the Gene*, 4th Edition, 1987, The Bejacmin/Cummings Pub. co., p.224). Conservative amino acid substitutions are made, for example, in accordance with those set forth in TABLE 1 as follows:

TABLE 1

	Original residue	Conservative substitution
	Ala (A)	Gly; Ser, Abu
20	Arg (R)	Lys, orn
	Asn (N)	Gln; His
	Cys (C)	Ser
	Gln (Q)	Asn
	Glu (E)	Asp
25	Gly (G)	Ala; Pro
	His (H)	Asn; Gln
	Ile (I)	Leu; Val; Met; Nle; Nva
	Leu (L)	Ile; Val; Met; Nle; Nva
	Lys (K)	Arg; Gln; Glu
30	Met (M)	Leu; Tyr; Ile; NLe Val
	Ornithine	Lys; Arg
	Phe (F)	Met; Leu; Tyr
	Ser (S)	Thr
	Thr (T)	Ser
35	Trp (W)	Tyr
	Tyr (Y)	Trp; Phe
	Val (V)	Ile; Leu; Met; Nle; Nva

-41-

Other substitutions are also permissible and may be determined empirically or in accord with known conservative substitutions.

As used herein, the amino acids, which occur in the various amino acid sequences appearing herein, are identified according to their well-known, three-letter or one-letter abbreviations. The nucleotides, which  
5 occur in the various DNA fragments, are designated with the standard single-letter designations used routinely in the art.

As used herein, a splice variant refers to a variant produced by differential processing of a primary transcript of genomic DNA that results  
10 in more than one type of mRNA.

As used herein, a probe or primer based on a nucleotide sequence includes at least 10, 14, 16, 30 or 100 contiguous nucleotides from the reference nucleic acid molecule.

As used herein, recombinant production by using recombinant DNA  
15 methods refers to the use of the well known methods of molecular biology for expressing proteins encoded by cloned DNA.

As used herein, biological activity refers to the *in vivo* activities of a compound or physiological responses that result upon *in vivo* administration of a compound, composition or other mixture. Biological  
20 activity, thus, encompasses therapeutic effects and pharmaceutical activity of such compounds, compositions and mixtures. Biological activities may be observed in *in vitro* systems designed to test or use such activities. Thus, for purposes herein the biological activity of a luciferase is its oxygenase activity whereby, upon oxidation of a  
25 substrate, light is produced.



-42-

The terms substantially identical or similar varies with the context as understood by those skilled in the relevant art and generally means at least 40, 60, 80, 90, 95 or 98%.

As used herein, substantially identical to a product means  
5 sufficiently similar so that the property is sufficiently unchanged so that the substantially identical product can be used in place of the product.

As used herein, substantially pure means sufficiently homogeneous to appear free of readily detectable impurities as determined by standard methods of analysis, such as thin layer chromatography (TLC), gel  
10 electrophoresis and high performance liquid chromatography (HPLC), used by those of skill in the art to assess such purity, or sufficiently pure such that further purification would not detectably alter the physical and chemical properties, such as enzymatic and biological activities, of the substance. Methods for purification of the compounds to produce  
15 substantially chemically pure compounds are known to those of skill in the art. A substantially chemically pure compound may, however, be a mixture of stereoisomers or isomers. In such instances, further purification might increase the specific activity of the compound.

As used herein, vector (or plasmid) refers to discrete elements that  
20 are used to introduce heterologous DNA into cells for either expression or replication thereof. The vectors typically remain episomal, but may be designed to effect integration of a gene or portion thereof into a chromosome of the genome. Also contemplated are vectors that are artificial chromosomes, such as yeast artificial chromosomes and  
25 mammalian artificial chromosomes. Selection and use of such vehicles are well known to those of skill in the art. An expression vector includes vectors capable of expressing DNA that is operatively linked with regulatory sequences, such as promoter regions, that are capable of effecting expression of such DNA fragments. Thus, an expression vector

-43-

refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector that, upon introduction into an appropriate host cell, results in expression of the cloned DNA.

Appropriate expression vectors are well known to those of skill in the art and include those that are replicable in eukaryotic cells and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome.

As used herein, protein-binding-sequence refers to a protein or peptide sequence that is capable of specific binding to other protein or peptide sequences generally, to a set of protein or peptide sequences or to a particular protein or peptide sequence.

As used herein, a composition refers to any mixture of two or more ingredients. It may be a solution, a suspension, liquid, powder, a paste, aqueous, non-aqueous or any combination thereof.

As used herein, a combination refers to any association between two or more items.

As used herein, fluid refers to any composition that can flow. Fluids thus encompass compositions that are in the form of semi-solids, pastes, solutions, aqueous mixtures, gels, lotions, creams and other such compositions.

As used herein, a cellular extract refers to a preparation or fraction that is made from a lysed or disrupted cell.

As used herein, the term "subject" refers to animals, plants, insects, and birds and other phyla, genera and species into which nucleic acid molecules may be introduced. Included are higher organisms, such as mammals, fish, insects and birds, including humans, primates, cattle, pigs, rabbits, goats, sheep, mice, rats, guinea pigs, hamsters, cats, dogs, horses, chicken and others.

-44-

As used herein, flow cytometry refers to processes that use a laser based instrument capable of analyzing and sorting out cells and or chromosomes based on size and fluorescence.

As used herein, the abbreviations for any protective groups, amino  
5 acids and other compounds, are, unless indicated otherwise, in accord with their common usage, recognized abbreviations, or the IUPAC-IUB Commission on Biochemical Nomenclature (see, (1972) *Biochem.* 11:942-944).

**B. Recombination systems**

10 Site-specific recombination systems typically contain three elements: a pair of DNA sequences (the site-specific recombination sequences) and a specific enzyme (the site-specific recombinase). The site-specific recombinase catalyzes a recombination reaction between two site-specific recombination sequences.

15 A number of different site-specific recombinase systems are available and/or known to those of skill in the art, including, but not limited to: the Cre/*lox* recombination system using CRE recombinase (see, *e.g.*, SEQ ID Nos. 58 and 59) from the Escherichia coli phage P1 (see, *e.g.*, Sauer (1993) *Methods in Enzymology* 225:890-900; Sauer *et al.* (1990) *The New Biologist* 2:441-449), Sauer (1994) *Current Opinion in Biotechnology* 5:521-527; Odell *et al.* (1990) *Mol Gen Genet.* 223:369-378; Lasko *et al.* (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:6232-6236; U.S. Patent No. 5,658,772), the FLP/FRT system of yeast using the FLP recombinase (see, SEQ ID Nos. 60 and 61) from the 2 $\mu$  episome of  
20 *Saccharomyces cerevisiae* (Cox (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80:4223; Falco *et al.* (1982) *Cell* 29:573-584; Golic *et al.* (1989) *Cell* 59:499-509; U.S. Patent No. 5,744,336), the resolvases, including Gin recombinase of phage Mu (Maeser *et al.* (1991) *Mol Gen Genet.* 230:170-176; Klippel, A. *et al* (1993) *EMBO J.* 12:1047-1057; see, *e.g.*,

-45-

SEQ ID Nos. 64-67), Cin, Hin,  $\alpha\delta$  Tn3; the Pin recombinase of *E. coli* (see, e.g., SEQ ID Nos. 68 and 69; Enomoto *et al.* (1983) *J. Bacteriol.* 6:663-668), the R/RS system of the pSR1 plasmid of *Zygosaccharomyces rouxii* (Araki *et al.* (1992) *J. Mol. Biol.* 225:25-37; Matsuzaki *et al.* (1990) *J. Bacteriol.* 172: 610-618) and site-specific recombinases from *Kluyveromyces drosophilarius* (Chen *et al.* (1986) *Nucleic Acids Res.* 314:4471-4481) and *Kluyveromyces waltii* (Chen *et al.* (1992) *J. Gen. Microbiol.* 138:337-345). Other systems are known to those of skill in the art (Stark *et al. Trends Genet.* 8:432-439; Utatsu *et al.* (1987) *J. Bacteriol.* 169:5537-5545; see, also, U.S. Patent No. 6,171,861).

Members of the highly related family of site-specific recombinases, the resolvase family, such as  $\gamma\delta$ , Tn3 resolvase, Hin, Gin, and Cin are also available. Members of this family of recombinases are typically constrained to intramolecular reactions (e.g., inversions and excisions) and can require host-encoded factors. Mutants have been isolated that relieve some of the requirements for host factors (Maeser *et al.* (1991) *Mol. Gen. Genet.* 230:170-176), as well as some of the constraints of intramolecular recombination (see, U.S. Patent No. 6,171,861).

The bacteriophage P1 Cre/lox and the yeast FLP/FRT systems are particularly useful systems for site-specific integration, inversion or excision of heterologous nucleic acid into, and out of, chromosomes, particularly ACes as provided herein. In these systems a recombinase (Cre or FLP) interacts specifically with its respective site-specific recombination sequence (lox or FRT, respectively) to invert or excise the intervening sequences. The sequence for each of these two systems is relatively short (34 bp for lox and 47 bp for FRT).

The FLP/FRT recombinase system has been demonstrated to function efficiently in plant cells (U.S. Patent No. 5,744,386), and, thus, can be used for producing plant artificial chromosome platforms. In

-46-

general, short incomplete FRT sites leads to higher accumulation of excision products than the complete full-length FRT sites. The system catalyzes intra- and intermolecular reactions, and, thus, can be used for DNA excision and integration reactions. The recombination reaction is  
5 reversible and this reversibility can compromise the efficiency of the reaction in each direction. Altering the structure of the site-specific recombination sequences is one approach to remedying this situation. The site-specific recombination sequence can be mutated in a manner that the product of the recombination reaction is no longer recognized as  
10 a substrate for the reverse reaction, thereby stabilizing the integration or excision event.

In the Cre-lox system, discovered in bacteriophage P1, recombination between loxP sites occurs in the presence of the Cre recombinase (see, *e.g.*, U.S. Patent No. 5,658,772). This system can be  
15 used to insert, invert or excise nucleic acid located between two lox sites. Cre can be expressed from a vector. Since the lox site is an asymmetrical nucleotide sequence, lox sites on the same DNA molecule can have the same or opposite orientation with respect to each other. Recombination between lox sites in the same orientation results in a deletion of the DNA  
20 segment located between the two lox sites and a connection between the resulting ends of the original DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule  
25 result in an inversion of the nucleotide sequence of the DNA segment located between the two lox sites. In addition, reciprocal exchange of DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the product of the Cre coding region.

-47-

Any site-specific recombinase system known to those of skill in the art is contemplated for use herein. It is contemplated that one or a plurality of sites that direct the recombination by the recombinase are introduced into an artificial chromosome to produce platform ACes. The  
5 resulting platform ACes are introduced into cells with nucleic acid encoding the cognate recombinase, typically on a vector, and nucleic acid encoding heterologous nucleic acid of interest linked to the appropriate recombination site for insertion into the platform ACes. The recombinase-encoding-nucleic acid may be introduced into the cells on the same  
10 vector, or a different vector, encoding the heterologous nucleic acid.

An *E. coli* phage lambda integrase system for ACes platform engineering and for artificial chromosome engineering is provided (Lorbach *et al.* (2000) *J. Mol. Biol* 296:1175-1181). The phage lambda integrase (Landy, A. (1989) *Annu. Rev. Biochem.* 58:913-94) is adapted herein and  
15 the cognate *att* sites are provided. Chromosomes, including ACes, engineered to contain one or a plurality of *att* sites are provided, as are vectors encoding a mutant integrase that functions in the absence other factors. Methods using the modified chromosomes and vectors for introduction of heterologous nucleic acid are also provided.

20 For purposes herein, one or more of the sites (e.g., a single site or a pair of sites) required for recombination are introduced into an artificial chromosome, such as an ACes chromosome. The enzyme for catalyzing site-directed recombination is introduced with the DNA of interest, or separately, or is engineered onto the artificial chromosome under the  
25 control of a regulatable promoter.

As described herein, artificial chromosome platforms containing one or multiple recombination sites are provided. The methods and resulting products are exemplified with the lambda phage Att/Int system, but

-48-

similar methods may be used for production of *ACes* platforms with other recombination systems.

The Att/Int system and vectors provided herein are not only intended for engineering *ACes* platforms, but may be used to engineer an Att/Int system into any chromosome. Introduction of att sites into a chromosome will permit engineering of natural chromosomes, such as by permitting targeted integration genes or regulatory regions, and by controlled excision of selected regions. For example, genes encoding a particular trait may be added to a chromosome, such as plant chromosome engineered to contain one or plurality of att sites. Such chromosomes may be used for screening DNA to identify genes. Large pieces of DNA can be introduced into cells and the cells screened phenotypically to select those having the desired trait.

#### C. Platforms

Provided herein are platform artificial chromosomes (platform *ACes*) containing single or multiple site-specific recombination sites. Chromosome-based platform technology permits efficient and tractable engineering and subsequent expression of multiple gene targets. Methods are provided that use DNA vectors and fragments to create platform artificial chromosomes, including animal, particularly mammalian, artificial chromosomes, and plant artificial chromosomes. The artificial chromosomes contain either single or multiple sequence-specific recombination sites suitable for the placement of target gene expression vectors onto the platform chromosome. The engineered chromosome-based platform *ACes* technology is applicable for methods, including cellular and transgenic protein production, transgenic plant and animal production and gene therapy. The platform *ACes* are also useful for producing a library of *ACes* comprising random portions of a given genome (e.g., a mammalian, plant or prokaryotic genome) for genomic

-49-

screening; as well as a library of cells comprising different and/or mutually exclusive *ACes* therein.

Exemplary of artificial chromosome platforms are those based on *ACes*. *ACes* artificial chromosomes are non-viral, self-replicating nucleic acid molecules that function as a natural chromosome, having all the elements required for normal chromosomal replication and maintenance within the cell nucleus. *ACes* artificial chromosomes do not rely on integration into the genome of the cell to be effective, and they are not limited by DNA carrying capacity and as such the therapeutic gene(s) of interest, including regulatory sequences, can be engineered into the *ACes*. In addition, *ACes* are stable *in vitro* and *in vivo* and can provide predictable long-term gene expression. Once engineered and delivered to the appropriate cell or embryo, *ACes* work independently alongside host chromosomes, for *ACes* that are predominantly heterochromatin producing only the products (proteins) from the genes it carries. As provided herein *ACes* are modified by introduction of recombination site(s) to provide a platform for ready introduction of heterologous nucleic acid. The *ACes* platforms can be used for production of transgenic animals and plants; as vectors for genetic therapy; for use as protein production systems; for animal models to identify and target new therapeutics; in cell culture for the development and production of therapeutic proteins; and for a variety of other applications.

#### 1. Generation of artificial chromosomes

Artificial chromosomes may be generated by any method known to those of skill in the art. Of particular interest herein are the *ACes* artificial chromosomes, which contain a repeated unit. Methods for production of *ACes* are described in detail in U.S. Patent Nos. 6,025,155 and 6,077,697, which, as with all patents, applications, publications and other disclosure, are incorporated herein in their entirety.



-50-

**Generation of *de novo* ACes.**

ACes can be generated by cotransfecting exogenous DNA—such as a mammary tissue specific DNA cassette including the gene sequences for a therapeutic protein, with a rDNA fragment and a drug resistance  
5 marker gene into the desired eukaryotic cell, such as plant or animal cells; such as murine cells *in vitro*. DNA with a selectable or detectable marker is introduced, and can be allowed to integrate randomly into pericentric heterochromatin or can be targeted to pericentric heterochromatin, such as that in rDNA gene arrays that reside on acrocentric chromosomes,  
10 such as the short arms of acrocentric chromosomes. This integration event activates the “megareplicator” sequence and amplifies the pericentric heterochromatin and the exogenous DNA, and duplicates a centromere. Ensuing breakage of this “dicentric” chromosome can result in the production of daughter cells that contain the substantially-original  
15 chromosome and the new artificial chromosome. The resulting ACes contain all the essential elements needed for stability and replication in dividing cells—centromere, origins of replications, and telomeres. ACes have been produced that express marker genes (lacZ, green fluorescent protein, neomycin-resistance, puromycin-resistance, hygromycin-  
20 resistance) and genes of interest. Isolated ACes, for example, have been successfully transferred intact to rodent, human, and bovine cells by electroporation, sonoporation, microinjection, and transfection with lipids and dendrimers.

To render the creation of ACes with desired genes more tractable  
25 and efficient, “platform” ACes (platform-ACes) can be produced that contain defined DNA sequences for enzyme-mediated homologous DNA recombination, such as by Cre or FLP recombinases (Bouhassira *et al.* (1996) *Blood* 88(supplement 1):190a; Bouhassira *et al.* (1997) *Blood*, 90:3332-3344; Siebler *et al.* (1997) *Biochemistry*: 36:1740-1747;

-51-

Siebler *et al.* (1998) *Biochemistry* 37: 6229-6234; and Bethke *et al.* (1997) *Nucl. Acids Res.* 25:2828-2834), and as exemplified herein the lambda phage integrase. A *lox* site contains two 13 bp inverted repeats to which Cre-recombinase binds and an intervening 8 bp core region.

- 5 Only pairs of sites having identity in the central 6 bp of the core region are proficient for recombination; sites having non-identical core sequences (heterospecific *lox* sites) do not efficiently recombine with each other (Hoess *et al.* (1986) *Nucleic Acids Res.* 14:2287-2300).

10 **Generating acrocentric chromosomes for plant artificial chromosome formation.**

In human and mouse cells *de novo* formation of a satellite DNA based artificial chromosome (SATAC, also referred to as ACes) can occur in an acrocentric chromosome where the short arm contains only pericentric heterochromatin, the rDNA array, and telomere sequences.

- 15 Plant species may not have any acrocentric chromosomes with the same physical structure described, but "megareplicator" DNA sequences reside in the plant rDNA arrays, also known as the nucleolar organizing regions (NOR). A structure like those seen in acrocentric mammalian chromosomes can be generated using site-specific recombination between  
20 appropriate arms of plant chromosomes.

**Approach**

- Qin *et al.* ((1994) *Proc. Natl. Acad. Sci. U.S.A.* 91:1706-1710, 1994) describes crossing two *Nicotiana tabacum* transgenic plants. One plant contains a construct encoding a promoterless hygromycin-resistance  
25 gene preceded by a *lox* site (*lox-hpt*), the other plant carries a construct containing a cauliflower mosaic virus 35S promoter linked to a *lox* sequence and the *cre* DNA recombinase coding region (35S-*lox-cre*). The constructs were introduced separately by infecting leaf explants with *agrobacterium tumefaciens* which carries the kanamycin-resistance gene

-52-

(Kan<sup>R</sup>). The resultant Kan<sup>R</sup> transgenic plants were crossed. Plants that carried the appropriate DNA recombination event were identified by hygromycin-resistance.

5                   **Modification of the above for generation of ACes**

The Kan<sup>R</sup> cultivars are initially screened, such as by FISH, to identify two sets of candidate transgenic plants. One set has one construct integrated in regions adjacent to the pericentric heterochromatin on the short arm of any chromosome. The second set of candidate plants  
10 has the other construct integrated in the NOR region of appropriate chromosomes. To obtain reciprocal translocation both sites must be in the same orientation. Therefore a series of crosses are required, Kan<sup>R</sup> plants generated, and FISH analyses performed to identify the appropriate "acrocentric" plant chromosome for *de novo* plant ACes formation.

15           **2. Bacteriophage lambda integrase-based site-specific recombination system**

An integral part of the platform technology includes a site-specific recombination system that allows the placement of selected gene targets or genomic fragments onto the platform chromosomes. Any such system  
20 may be used. In particular, a method is provided for insertion of additional DNA fragments into the platform chromosome residing in the cell via sequence-specific recombination using the recombinase activity of the bacteriophage lambda integrase. The lambda integrase system is exemplary of the recombination systems contemplated for ACes. Any  
25 known recombination system, including any described herein, particularly any that operates without the need for additional factors or that, by virtue of mutation, does not require additional factors, is contemplated.

-53-

As noted the lambda integrase system provided herein can be used with natural chromosomes and artificial chromosomes in addition to *ACes*. Single or a plurality of recombination sites, which may be the same or different, are introduced into artificial chromosomes to produce  
5 artificial chromosome platforms.

**3. Creation of bacteriophage lambda integrase site-specific recombination system**

The lambda phage-encoded integrase (designated *Int*) is a prototypical member of the integrase family. *Int* effects integration and  
10 excision of the phage in and out of the *E. coli* genome via recombination between pairs of attachment sites designated *attB/attP* and *attL/attR*. Each *att* site contains two inverted 9 base pair core *Int* binding sites and a 7 base pair overlap region that is identical in wild-type *att* sites. Each  
15 site, except for *attB* contains additional *Int* binding sites. In flanking regions, there are recognition sequences for accessory DNA binding proteins, such as integration host factor (IHF), factor for inversion stimulation (FIS) and the phage encoded excision protein (XIS). Except  
20 for *attB*, *Int* is a heterobivalent DNA-binding protein and, with assistance from the accessory proteins and negative DNA supercoiling, binds simultaneously to core and arm sites within the same *att* site.

*Int*, like Cre and FLP, executes an ordered sequential pair of strand exchanges during integrative and excisive recombination. The natural pairs of target sequences for *Int*, *attB* and *attP* or *attL* and *attR* are located on the same or different DNA molecules resulting in intra or  
25 intermolecular recombination, respectively. For example, intramolecular recombination occurs between inversely oriented *attB* and *attP*, or between *attL* and *attR* sequences, respectively, leading to inversion of the intervening DNA segment.

-54-

Like the recombinase systems, such as Cre and FLP, Int directs site-specific recombination. Unlike the other systems, such as Cre and FLP, Int generally requires additional protein factors for integrative and excisive recombination and negative supercoiling for integrative recombination.

5 Hence, the Int system had not been used in eukaryotic targeting systems.

Mutant Int proteins, designated Int-h (E174K) and a derivative thereof Int-h/218(E174K/E218K) do not require accessory proteins to perform intramolecular integrative and excisive recombination in co-transfection assays in human cells (Lorbach *et al.* (2000) *J Mol. Biol.* 10 296:1175-1181); wild-type Int does not catalyze intramolecular recombination in human cells harboring target sites *attB* and *attP*. Hence it had been demonstrated that mutant Int can catalyze factor-independent recombination events in human cells.

There has been no demonstration by others that this system can be 15 used for engineering of eukaryotic genomes or chromosomes. Provided herein are chromosomes, including artificial chromosomes, such as but not limited to ACes that contain *att* sites (e.g., platform ACes), and the use of such chromosomes for targeted integration of heterologous DNA into such chromosomes in eukaryotic cells, including animal, such as 20 rodent and human, and plant cells. Mutant Int provided herein is shown to effect site-directed recombination between sites in artificial chromosomes and vectors containing cognate sites.

An additional component of the chromosome-based platform technology is the site-specific integration of target DNA sequences onto 25 the platform. For this the native bacteriophage lambda integrase has been modified to carry out this sequence specific DNA recombination event in eukaryotic cells. The bacteriophage lambda integrase and its cognate DNA substrate *att* is a member of the site-specific recombinase family that also includes the bacteriophage P1 Cre/lox system as well as

-55-

the *Saccharomyces cerevisiae* 2 micron based FLP/FRT system (see, e.g., Landy (1989) *Ann. Rev. Biochem* 58:913-949; Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398-3402; Broach *et al.* (1982) *Cell* 29:227-234).

5 By combining DNA endonuclease and DNA ligase activity these recombinases recognize and catalyze DNA exchanges between sequences flanking the recognition site. During the integration of lambda genome into the *E. coli* (lambda recombination) genome, the phage integrase (INT) in association with accessory proteins catalyzes the DNA exchange  
10 between the attP site of the phage genome and the attB site of the bacterial genome resulting in the formation of attL and attR sites (Figure 6). The engineered bacteriophage lambda integrase has been produced herein to carry out an intermolecular DNA recombination event between an incoming DNA molecule (primarily on a vector containing the bacterial  
15 attB site) and the chromosome-based platform carrying the lambda attP sequence independent of lambda bacteriophage or bacterial accessory proteins.

In contrast to the bi-directional Cre/lox and FLP/FRT system, the engineered lambda recombination system derived for chromosome-based  
20 platform technology is advantageously unidirectional because accessory proteins, which are absent, are required for excision of integrated nucleic acid upon further exposure to the lambda Int recombinase.

**4. Creation of platform chromosome containing single or multiple sequence-specific recombination sites**

**25 a. Multiple sites**

For the creation of a platform chromosome containing multiple, sequence-specific recombination sites, artificial chromosomes are produced as depicted in Figure 5 and Example 3. As discussed above, artificial chromosomes can be produced using any suitable methodology,

-56-

including those described in U.S. Patent Nos. 5,288,625; 5,712,134; 5,891,691; 6,025,155. Briefly, to prepare artificial chromosomes containing multiple recombination (e.g., integration) sites, nucleic acid (either in the form a one or more plasmids, such as the plasmid pSV40193attPsensePUR set forth in Example 3) is targeted into an amplifiable region of a chromosome, such as the pericentric region of a chromosome. Among such regions are the rDNA gene loci in acrocentric mammalian chromosomes. Hence, targeting nucleic acid for integration into the rDNA region of mammalian acrocentric chromosomes can include the mouse rDNA fragments (for targeting into rodent cell lines) or large human rDNA regions on BAC/PAC vectors (or subclones thereof in standard vectors) for targeting into human acrocentric chromosomes, such as for human gene therapy applications. The targeting nucleic acid generally includes a detectable or selectable marker, such as antibiotic resistance, such as puromycin and hygromycin, a recombination site (such as attP, attB, attL, attR or the like), and/or human selectable markers as required for gene therapy applications. Cells are grown under conditions that result in amplification and ultimately production of ACes artificial chromosomes having multiple recombination (e.g., integration) sites therein. ACes having the desired size are selected for further engineering.

**b. Creation of platform chromosome containing a single sequence-specific recombination site**

In this method a mammalian platform artificial chromosome is generated containing a single sequence-specific recombination site. In the Example below, this approach is demonstrated using a puromycin resistance marker for selection and a mouse rDNA fragment for targeting into the rDNA locus on mouse acrocentric chromosomes. Other selection markers and targeting DNA sequences as desired and known to those of

-57-

skill in the art can be used. Additional resistance markers include genes conferring resistance to the antibiotics neomycin, blasticidin, hygromycin and zeocin. For applications, such as gene therapy in which potentially immunogenic responses are to be avoided, host, such as human, derived

5 selectable markers or markers detectable with monoclonal antibodies (MAb) followed by fluorescent activated cell sorting (FACS) can be used. Examples in this class include, but are not limited to: human nerve growth factor receptor (detection with MAb); truncated human growth factor receptor (detection with MAb); mutant human dihydrofolate reductase

10 (DHFR; detectable using a fluorescent methotrexate substrate); secreted alkaline phosphatase (SEAP; detectable with fluorescent substrate); thymidylate synthase (TS; confers resistance to fluorodeoxyuridine); human CAD gene (confers resistance to N-phosphonacetyl-L-aspartate (PALA)).

15 To construct a platform artificial chromosome with a single site, an ACes artificial chromosome (or other artificial chromosome of interest) can be produced containing a selectable marker. A single sequence specific recombination site is targeted onto ACes via homologous recombination. For this, DNA sequences containing the site-specific

20 recombination sequence are flanked with DNA sequences homologous to a selected sequence in the chromosome. For example, when using a chromosome containing rDNA or satellite DNA, such DNA can be used as homologous sequences to target the site-specific recombination sequence onto the chromosome. A vector is designed to have these homologous

25 sequences flanking the site-specific recombination site and, after the appropriate restriction enzyme digest to generate free ends of homology to the chromosome, the DNA is transfected into cells harboring the chromosome. After transfection and integration of the site-specific cassette, homologous recombination events onto the platform



-58-

chromosome are subcloned and identified, for example by screening single cell subclones via expression of resistance or a fluorescent marker and PCR analysis. In one embodiment, a platform artificial chromosome, such as a platform ACes, that contains a single copy of the recombination site is selected. Examples 2B and 2D exemplify the process, and Figure 3 provides a diagram depicting one method for the creation of a platform mammalian chromosome containing a single sequence-specific recombination site.

5. **Lambda integrase mediated recombination of target gene expression vector onto platform chromosome**

The third component of the chromosome-based platform technology involves the use of target gene expression vectors carrying, for example, genes for gene therapy, genes for transgenic animal or plant production, and those required for cellular protein production of interest. Using lambda integrase mediated site-specific recombination, or any other recombinase-mediated site-specific recombination, the target gene expression vectors are introduced onto the selected chromosome platform. The use of target gene expression vector permits use of the *de novo* generated chromosome-based platforms for a wide range of gene targets. Furthermore, chromosome platforms containing multiple *attP* sites provides the opportunity to incorporate multiple gene targets onto a single platform, thereby providing for expression of multiple gene targets, including the expression of cellular and genetic regulatory genes and the expression of all or parts of metabolic pathways. In addition to expressing small target genes, such as cDNA and hybrid cDNA/artificial intron constructs, the chromosome-based platform can be used for engineering and expressing large genomic fragments carrying target genes along with its endogenous genomic promoter sequences. This is of importance, for example, where the therapy requires precise cell specific

-59-

expression and in instances where expression is best achieved from genomic clones rather than cDNA clones. Figure 9 provides a diagram summarizing one embodiment of the chromosome-based technology.

A feature of the target gene expression vector that is of interest to  
5 include is a promoterless marker gene, which as exemplified (see, Figure 9) contains an upstream *attB* site (marker 2 on Figure 9). The nucleic acid encoding the marker is not expressed unless it is placed downstream from a promoter sequence. Using the recombinase technology provided herein, such as the lambda integrase technology ( $\lambda$ INT<sub>E174R</sub> on figure 8)  
10 provided herein, site-specific recombination between the *attB* site on the vector and the promoter-*attP* site (in the "sense" orientation) on the chromosome-based platform results in the expression of marker 2 on the target gene expression vector, thereby providing a positive selection for the lambda INT mediated site-specific recombination event. Site-specific  
15 recombination events on the chromosome-based platform versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR. Examples of suitable marker 2 genes, include, but are not limited to, genes that confer resistance to toxic compounds or antibiotics,  
20 fluorescence activated cell sorting (FACS) sortable cell surface markers and various fluorescent markers. Examples of these genes include, but are not limited to, human L26a<sup>R</sup> (human homolog of *Saccharomyces cerevisiae* CYH<sup>B</sup> gene), neomycin, puromycin, blasticidin, CD24 (see, e.g., US Patents 5,804,177 and 6,074,836), truncated CD4, truncated low  
25 affinity nerve growth factor receptor (LNGFR), truncated LDL receptor, truncated human growth hormone receptor, GFP, RFP, BFP.

The target gene expression vectors contain a gene (target gene) for expression from the chromosome platform. The target gene can be expressed using various constitutive or regulated promoter systems

-60-

across various mammalian species. For the expression of multiple target genes within the same target gene expression vector, the expression of the multiple targets can be coordinately regulated via viral-based or human internal ribosome entry site (IRES) elements (see, *e.g.*, Jackson *et al.* (1990) *Trends Biochem Sci.* 15: 477-83; Oumard *et al.* (2000) *Mol. Cell. Biol.* 20: 2755-2759). Furthermore, using IRES type elements linked to a downstream fluorescent marker, *e.g.*, green, red or blue fluorescent proteins (GFP, RFP, BFP) allows for the identification of high expressing clones from the integrated target gene expression vector.

- 10 In certain embodiments described herein, the promoterless marker can be transcriptionally downstream of the heterologous nucleic acid, wherein the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.
- 15 In addition, the selectable marker and the heterologous nucleic acid can be transcriptionally linked by the presence of a IRES between them. As set forth herein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.
- 20 Expression from the target gene expression vector integrated onto the chromosome-based platform can be further enhanced using genomic insulator/boundary elements. The incorporation of insulator sequences into the target gene expression vector helps define boundaries in chromatin structure and thus minimizes influence of chromatin position effects/gene silencing on the expression of the target gene (Bell *et al.* (1999) *Current Opinion in Genetics and Development* 9:191-198; Emery *et al.* (2000) *Proc. Natl. Acad. Sci. U.S.A.* 97:9150-9155). Examples of insulator elements that can be included onto target gene expression vector in order to optimize expression include, but are not limited to:
- 25

-61-

- 1) chicken  $\beta$ -globin HS4 element (Prioleau *et al.* (1999) *EMBO J* 18: 4035-4048);
- 2) matrix attachment regions (MAR; see, *e.g.*, Ramakrishnan *et al.* (2000) *Mol Cell. Biol.* 20:868-877);
- 5 3) scaffold attachment regions (SAR; see, *e.g.*, Auten *et al.* (1999) *Human Gene Therapy* 10:1389-1399); and
- 4) universal chromatin opening elements (UCOE; WO/0005393 and WO/0224930)

The copy number of the target gene can be controlled by sequentially adding multiple target gene expression vectors containing the target gene onto multiple integration sites on the chromosome platform. Likewise, the copy number of the target gene can be controlled within an individual target gene expression vector by the addition of DNA sequences that promote gene amplification. For example, gene amplification can be induced utilizing the dihydrofolate reductase (DHFR) minigene with subsequent selection with methotrexate (see, *e.g.*, Schimke (1984) *Cell* 37:705-713) or amplification promoting sequences from the rDNA locus (see, *e.g.*, Wegner *et al.* (1989) *Nucl. Acids Res.* 17: 9909-9932).

## 20 6. Platforms with other recombinase system sites

A "double *lox*" targeting strategy mediated by Cre-recombinase (Bethke *et al.* (1997) *Nucl. Acids Res.* 25:2828-2834) can be used. This strategy employs a pair of heterospecific *lox* sites—*loxA* and *loxB*, which differ by one nucleotide in the 8 bp spacer region. Both sites are engineered into the artificial chromosome and also onto the targeting DNA vector. This allows for a direct site-specific insertion of a commercially relevant gene or genes by a Cre-catalyzed double crossover event. In essence a platform ACes is engineered with a hygromycin-resistance gene flanked by the double *lox* sites generating *lox-ACes*, which is maintained

-62-

in the thymidine kinase deficient cell, LMtk(-). The gene of interest, for example, for testing purposes, the green fluorescence protein gene, GFP and a HSV thymidine kinase gene (tk) marker, are engineered between the appropriate *lox* sites of the targeting vector. The vector DNA is

5 cotransfected with plasmid pBS185 (Life Technologies) encoding the *Cre* recombinase gene into mammalian cells maintaining the dual-*lox* artificial chromosome. Transient expression of the *Cre* recombinase catalyzes the site-specific insertion of the gene and the tk-gene onto the artificial chromosome. The transfected cells are grown in HAT medium that

10 selects for only those cells that have integrated and expressed the thymidine kinase gene. The HAT<sup>R</sup> colonies are screened by PCR analyses to identify artificial chromosomes with the desired insertion.

To generate the *lox-ACes*, Lambda-Hyg<sup>R</sup>-*lox* DNA is transfected into the LMtk(-) cell line harboring the precursor *ACes*. Hygromycin-

15 resistant colonies are analyzed by FISH and Southern blotting for the presence of a single copy insert on the *ACes*.

To demonstrate the gene replacement technology, cell lines containing candidate *lox-ACes* are cotransfected with pTK-GFP-*lox* and pBS185 (encoding the *Cre* recombinase gene) DNA. After transfection,

20 transient expression of plasmid pBS185 will provide sufficient burst of *Cre* recombinase activity to catalyze DNA recombination at the *lox* sites. Thus, a double crossover event between the *ACes* target and the exogenous targeting plasmid carrying the *loxA* and *loxB* permits the simple replacement of the hygromycin-resistance gene on the *lox-ACes*

25 for the tk-GFP cassette from the targeting plasmid, with no integration of vector DNA. Transfected cells are grown in HAT-media to select for tk-expression. Correct targeting will result in the generation of HAT<sup>R</sup>, hygromycin sensitive, and green fluorescent cells. The desired integration event is verified by Southern and PCR analyses. Specific PCR primer sets

-63-

are used to amplify DNA sequences flanking the individual *loxA* and *loxB* sites on the *lox-ACes* before and after homologous recombination.

**D. Exemplary applications of the Platform *ACes***

Platform *ACes* are applicable and tractable for different/optimized  
5 cell lines. Those that include a fluorescent marker, for example, can be  
purified and isolated using fluorescent activated cell sorting (FACS), and  
subsequently delivered to a target cell. Those with selectable markers  
provide for efficient selection and provide a growth advantage. Platform  
*ACes* allow multiple payload delivery of donor target vectors via a  
10 positive-selection site-specific, recombination system, and they allow for  
the inclusion of additional genetic factors that improve protein production  
and protein quality.

The construction and use of the platform *ACes* as provided for  
each application may be similarly applied to other applications. Particular  
15 descriptions are for exemplification.

**1. Cellular Protein Production Platform *ACes* (CPP *ACes*)**

As described herein, *ACes* can be produced from acrocentric  
chromosomes in rodent (mouse, hamster) cell lines via megareplicator  
induced amplification of heterochromatin/rDNA sequences. Such *ACes*  
20 are ideal for cellular protein production as well as other applications  
described herein and known to those of skill in the art. *ACes* platforms  
that contain a plurality of recombination sites are particularly suitable for  
engineering as cellular protein production systems.

In one embodiment, CPP *ACes* involve a two-component system:  
25 the platform chromosome containing multiple engineering sites and the  
donor target vector containing a platform-specific recombination site with  
designed expression cassettes (see Figure 9).

The platform *ACes* can be produced from any artificial  
chromosome, particularly the amplification-based artificial chromosomes.

-64-

For exemplification, they are produced from rodent artificial chromosomes produced from acrocentric chromosomes using the technology of U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183, in which nucleic acid is targeted to the  
5 pericentric heterochromatic, and, particularly into rDNA to initiate the replication event(s). The ACes can be produced directly in the chosen cellular protein production cell lines, such as, but not limited to, CHO cells, hybridomas, plant cells, plant tissues, plant protoplasts, stem cells and plant calli.

10                   a.       **Platform Construction**

In the exemplary embodiment, the initial *de novo* platform construction requires co-transfecting with excess targeting DNA, such as, rDNA or lambda DNA without an *attP* region, and an engineered selectable marker. The engineered selectable marker should contain  
15 promoter, generally a constitutive promoter, such as human, viral, i.e., adenovirus or SV40 promoter, including the human ferritin heavy chain promoter (SEQ ID NO:128), SV40 and EF1 $\alpha$  promoters, to control expression of a marker gene that provides a selective growth advantage to the cell. An example of such a marker gene is the *E. coli hisD* gene  
20 (encoding histidinol dehydrogenase) which is homologous and analogous to the *S. typhimurium hisD* a dominant marker selection system for mammalian cells previously described (see, Hartman *et al.* (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85:8047-8051). Since histidine is an essential amino acid in mammals and a nutritional requirement in cell culture, the *E.*  
25 *coli hisD* gene can be used to select for histidine prototrophy in defined media. Furthermore more stringent selection can be placed on the cells by including histinol in the medium. Histidinol is itself permeable and toxic to cells. The *hisD* provides a means of detoxification.

-65-

Placed between the promoter and the marker gene is the bacteriophage lambda *attP* site to use the bacteriophage lambda integrase dependent site-specific recombination system (described herein). The insertion of an *attP* site downstream of a promoter element provide  
5 forward selection of site-specific recombination events onto the platform *ACes*.

**b. Donor Target Vector Construction**

A second component of the CPP platform *ACes* system involves the construction of donor target vectors containing a gene product(s) of  
10 interest for the CPP platform *ACes*. Individual donor target vectors can be designed for each gene product to be expressed thus enabling maximum usage of a *de novo* constructed platform *ACes*, so that one or a few CPP platform *ACes* will be required for many gene targets.

A key feature of the donor vector target is the *promoterless* marker  
15 gene containing an upstream *attB* site (marker 2 on figure 9). Normally the marker would not be expressed unless it is placed downstream of a promoter sequence. As discussed above, using the lambda integrase technology ( $\lambda$ INT<sub>E174R</sub> on Figure 8 and Figure 9), site-specific recombination between the *attB* site on the vector and the promoter-*attP*  
20 site on the CPP platform *ACes* result in the expression of the donor target vector marker providing positive selection for the site-specific event. Site-specific recombination events on the CPP *ACes* versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR.  
25 In addition, since the lambda integrase reaction is unidirectional, i.e. excision reaction is not possible, a number of unique targets can be loaded onto the CPP platform *ACes* limited only by the number of markers available.



-66-

Additional features of the donor target vector include gene target expression cassettes flanked by either chromatin insulator regions, matrix attachment regions (MAR) or scaffold attachment regions (SAR). The use of these regions will provide a more "open" chromatin environment for gene expression and help alleviate silencing. An example of such a cassette for expressing a monoclonal antibody is described. For this purpose, a strong constitutive promoter, e.g. chicken  $\beta$ -actin or RNA Poll, is used to drive the expression of the heavy and light chain open reading frames. The heavy and light chain sequences flank a nonattenuated human IRES (IRES<sub>H</sub>; from the 5'UTR of NRF1 gene; see Oumard et al., 2000, *Mol. and Cell Biol.*, 20(8):2755-2759) element thereby coordinating transcription of both heavy and light chain sequence. Distal to the light chain open reading frame resides an additional viral encoded IRES (IRES<sub>V</sub> modified ECMV internal ribosomal entry site (IRES)) element attenuating the expression of the fluorescent marker gene hrGFP from *Renilla* (Stratagene). By linking the hrGFP with an attenuated IRES, the heavy and light chains along with the hrGFP are monocistronic. Thus, the identification of hrGFP fluorescing cells will provide a means to detect protein producing cells. In addition, high producing cell lines can be identified and isolated by FACS thereby decreasing the time frame in finding high expressers. Functional monoclonal antibody will be confirmed by ELISA.

**c. Additional components in cellular protein production platform ACes (CPP ACes)**

In addition to the aforementioned CPP ACes system, other genetic factors can be included to enhance the yield and quality of the expressed protein. Again to provide maximum flexibility, these additional factors can be inserted onto the CPP platform ACes by  $\lambda$ INTE174R dependent site-specific recombination. Other factors that could be used with a CPP

-67-

Platform *ACes* include for example, adenovirus E1a transactivation system which upregulates both cellular and viral promoters (see, e.g., Svensson and Akusjarvi (1984) EMBO 3:789-794; and US patents 5,866,359; 4,775,630 and 4,920,211).

5                   d.       **Targets for CHO-*ACes* engineering to enhance cell growth, such as CHO cell growth and protein production/ quality**

If adding these additional factors onto the CPP *ACes* is not prudent or desired, the host cell, CHO cells, can be engineered to express these  
10 factors (see, below, targets for CHO-*ACes* engineering to enhance CHO cell growth and protein production/quality). Additional factors to consider including are addition of insulin or IGF-1 to sustain viability;  
human sialyltransferases or related factors to produce more human-like glycoproteins; expression of factors to decrease ammonium accumulation  
15 during cell growth; expression of factors to inhibit apoptosis; expression of factors to improve protein secretion and protein folding; and expression of factors to permit serum-free transfection and selection.

1)       **Addition of insulin or IGF-1 to sustain viability**

20           Stimulatory factors and/or their receptors are expressed to set up an autocrine loop, to improve cell growth, such as CHO cell growth. Two exemplary candidates are insulin and IGF-1 (see, Biotechnol Prog 2000 Sep;16(5):693-7). Insulin is the most commonly used growth factor for sustaining cell growth and viability in serum-free Chinese hamster ovary  
25 (CHO) cell cultures. Insulin and IGF-1 analog (LongR(3) serve as growth and viability factors for CHO cells.

CHO cells were modified to produce higher levels of essential nutrients and factors. A serum-free (SF) medium for dihydrofolate reductase-deficient Chinese hamster ovary cells (DG44 cells) was  
30 prepared. Chinese hamster ovary cells (DG44 cells), which are normally

-68-

maintained in 10% serum medium, were gradually weaned to 0.5% serum medium to increase the probability of successful growth in SF medium (see, Kim *et al.* (199) *In Vitro Cell Dev Biol Anim* 35(4):178-82). A SF medium (SF-DG44) was formulated by supplementing the basal medium with these components; basal medium was prepared by supplementing Dulbecco's modified Eagle's medium and Ham's nutrient mixture F12 with hypoxanthine (10 mg/l) and thymidine (10 mg/l). Development of a SF medium for DG44 cells was facilitated using a Plackett-Burman design technique and weaning of cells.

10

**2) Human sialyltransferases or related factors to produce more human-like glycoproteins**

CHO cells have been modified by increasing their ability to process protein via addition of complex carbohydrates. This has been achieved by overexpression of relevant processing enzymes, or in some cases, reducing expression of relevant enzymes (see, Bragonzi *et al.* (2000) *Biochim Biophys Acta* 1474(3):273-282; see, also Weikert *et al.* (1999) *Nature biotech.* 17:1116-11121; Ferrari J *et al.* (1998) *Biotechnol Bioeng* 60(5):589-95). A CHO cell line expressing alpha2,6-sialyltransferase was developed for the production of human-like sialylated recombinant glycoproteins. The sialylation defect of CHO cells can be corrected by transfecting the alpha2,6-sialyltransferase (alpha2,6-ST) cDNA into the cells. Glycoproteins produced by such CHO cells display alpha2,6- and alpha2,3-linked terminal sialic acid residues, similar to human glycoproteins.

As another example for improving the production of human-like sialylated recombinant glycoproteins, a CHO cell line has been developed that constitutively expresses sialidase antisense RNA (see, Ferrari J *et al.* (1998) *Biotechnol Bioeng* 60(5):589-95). Several antisense expression

30

-69-

vectors were prepared using different regions of the sialidase gene. Co-transfection of the antisense constructs with a vector conferring puromycin resistance gave rise to over 40 puromycin resistant clones that were screened for sialidase activity. A 5' 474 bp coding segment of the  
5 sialidase cDNA, in the inverted orientation in an SV 40-based expression vector, gave maximal reduction of the sialidase activity to about 40% wild-type values.

Oligosaccharide biosynthesis pathways in mammalian cells have been engineered for generation of recombinant glycoproteins (see, *e.g.*,  
10 Sburlati (1998) *Biotechnol Prog* 14(2):189-92), which describes a Chinese hamster ovary (CHO) cell line capable of producing bisected oligosaccharides on glycoproteins. This cell line was created by overexpression of a recombinant N-acetylglucosaminyltransferase III (GnT-III) (see, also, Prati *et al.* (1998) *Biotechnol Bioeng* 59(4):445-50, which  
15 describes antisense strategies for glycosylation engineering of CHO cells).

**3) Expression of factors to decrease ammonium accumulation during cell growth**

Excess ammonium, which is a by-product of CHO cell metabolism  
20 can have detrimental effects on cell growth and protein quality (see, Yang *et al.* (2000) *Biotechnol Bioeng* 68(4):370-80). To solve this problem ammonium levels were modified by overexpressing carbamoyl phosphate synthetase I and ornithine transcarbamoylase or glutamine synthetase in CHO cells. Such modification resulted in reduced ammonium levels  
25 observed and an increase in the growth rate (see Kim *et al.* (2000) *J Biotechnol* 81(2-3):129-40; and Enosawa *et al.* (1997) *Cell Transplant* 6(5):537-40).

**4) Expression of factors to improve protein secretion and protein folding**

-70-

Overexpression of relevant enzymes can be engineered into the ACes to improve protein secretion and folding.

**5) Expression of factors to permit serum-free transfection and selection**

5 It is advantageous to have the ability to convert CHO cells in suspension growing in serum free medium to adherence with out having to resort to serum addition. Laminin or fibronectin addition is sufficient to make cells adherent (see, *e.g.*, Zaworski *et al.* (1993) *Biotechniques* 15(5):863-6) so that expressing either of these genes in CHO cells under  
10 an inducible promoter should allow for reversible shift to adherence without requiring serum addition.

**2. Platform ACes and Gene Therapy**

The platform ACes provided herein are contemplated for use in mammalian gene therapy, particularly human gene therapy. Human ACes  
15 can be derived from human acrocentric chromosomes from human host cells, in which the amplified sequences are heterochromatic and/or human rDNA. Different platform ACes applicable for different tissue cell types are provided. The ACes for gene therapy can contain a single copy of a therapeutic gene inserted into a defined location on platform ACes.  
20 Therapeutic genes include genomic clones, cDNA, hybrid genes and other combinations of sequences. Preferred selectable markers are those from the mammalian host, such as human derived factors so that they are non-immunogenic, non-toxic and allow for efficient selection, such as by FACS and/or drug resistance.

25 Platform ACes, useful for gene therapy and other applications, as noted herein, can be generated by megareplicator dependent amplification, such as by the methods in U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183. In one embodiment, human ACes are produced using

-71-

human rDNA constructs that target rDNA arrays on human acrocentric chromosomes and induce the megareplicator in human cells, particularly in primary cell lines (with sufficient number of doublings to form the ACes) or stem cells (such as hematopoietic stem cells, mesenchymal stem cells, adult stem cells or embryonic stem cells) to avoid the introduction of potentially harmful rearranged DNA sequences present in many transformed cell lines. Megareplicator induced ACes formation can result in multiple copies of targeting DNA/selectable markers in each amplification block on both chromosomal arms of the platform ACes.

10 In view of the considerations regarding immunogenicity and toxicity, the production of human platform ACes for gene therapy applications employs a two component system analogous to the platform ACes designed for cellular protein production (CPP platform ACes). The system includes a platform chromosome of entirely human DNA origin  
15 containing multiple engineering sites and a gene target vector carrying the therapeutic gene of interest.

**a. Platform Construction**

The initial *de novo* construction of the platform chromosome employs the co-transfection of excess targeting DNA and a selectable  
20 marker. In one embodiment, the DNA is targeted to the rDNA arrays on the human acrocentric chromosomes (chromosomes 13, 14, 15, 21 and 22). For example, two large human rDNA containing PAC clones 18714 and 18720 and the human PAC clone 558F8 are used for targeting (Genome Research (ML) now Incyte, BACPAC Resources, 747 52nd  
25 Street, Oakland CA). The mouse rDNA clone pFK161 (SEQ ID NO: 118), which was used to make the human SATAC from the 94-3 hamster/human hybrid cell line (see, *e.g.*, published International PCT application No. WO 97/40183 and Csonka, *et al*, *Journal of Cell Science*

-72-

113:3207-32161 and Example 1 for a description of pFK161) can also be used.

For animal applications, selectable markers should be non-immunogenic in the animal, such as a human, and include, but are not limited to: human nerve growth factor receptor (detected with a MAb, such as described in US patent 6,365,373); truncated human growth factor receptor (detected with MAb), mutant human dihydrofolate reductase (DHFR; fluorescent MTX substrate available); secreted alkaline phosphatase (SEAP; fluorescent substrate available); human thymidylate synthase (TS; confers resistance to anti-cancer agent fluorodeoxyuridine); human glutathione S-transferase alpha (GSTA1; conjugates glutathione to the stem cell selective alkylator busulfan; chemoprotective selectable marker in CD34+ cells); CD24 cell surface antigen in hematopoietic stem cells; human CAD gene to confer resistance to N-phosphonacetyl-L-aspartate (PALA); human multi-drug resistance-1 (MDR-1; P-glycoprotein surface protein selectable by increased drug resistance or enriched by FACS); human CD25 (IL-2 $\alpha$ ; detectable by Mab-FITC); Methylguanine-DNA methyltransferase (MGMT; selectable by carmustine); and Cytidine deaminase (CD; selectable by Ara-C).

Since megareplicator induced amplification generates multiple copies of the selectable marker, a second consideration for the selection of the human marker is the resulting dose of the expressed marker after ACes formation. High level of expression of certain markers may be detrimental to the cell and/or result in autoimmunity. One method to decrease the dose of the marker protein is by shortening its half-life, such as via the fusion of the well-conserved human ubiquitin tag (a 76 amino acid sequence) thus leading to increased turnover of the selectable marker. This has been used successfully for a number of reporter

-73-

systems including DHFR (see, *e.g.*, Stack *et al.* (2000) *Nature Biotechnology* 18:1298-1302 and references cited therein).

Using the ubiquitin tagged protein, a human selectable marker system analogous to the CPP ACes described herein is constructed.

- 5 Briefly, a tagged selectable marker, such as for example one of those described herein, is cloned downstream of an *attP* site and expressed from a human promoter. Exemplary promoters contemplated for use herein include, but are not limited to, the human ferritin heavy chain promoter (SEQ ID NO:128); RNA Poll; EF1 $\alpha$ ; TR; glyceraldehyde-3-  
10 phosphate dehydrogenase core promoter (GAP); a GAP core promoter including a proximal insulin inducible element the intervening GAP sequence; phosphofructokinase promoter; and phosphoglycerate kinase promoter. Also contemplated herein is an aldolase A promoter H1 & H2 (representing closely spaced transcriptional start sites) along with the  
15 proximal H enhancer. There are 4 promoters (*e.g.*, transcriptional start sites) for this gene, each having different regulatory and tissue activity. The H (most proximal 2) promoters are ubiquitously expressed off the H enhancer. This resulting marker can then be co-transfected along with excess human rDNA targeting sequence into the host cells. An important  
20 criteria for the selection of the

- recipient cells is sufficient number of cell doublings for the formation and detection of ACes. Accordingly, the co-transfections should be attempted in human primary cells that can be cultured for long periods of  
25 time, such as for example, stem cells (*e.g.*, hematopoietic, mesenchymal, adult or embryonic stem cells), or the like. Additional cell types, include, but are not limited to: single gene transfected cells exhibiting increased life-span; over-expressing c-myc cells, *e.g.* MSU1.1 (Morgan *et al.*, 1991, *Exp. Cell Res.*, Nov;197(1):125-136); over-expressing telomerase lines;



-74-

such as TERT cells; SV40 large T-antigen transfected lines; tumor cell lines, such as HT1080; and hybrid human cell lines, such as the 94-3 hamster/human hybrid cell line.

**b. Gene Target Vector**

5           The second component of the GT platform *ACes* (GT *ACes*) system involves the use of engineered target vectors carrying the therapeutic gene of interest. These are introduced onto the GT platform *ACes* via site-specific recombination. As with the CPP *ACes*, the use of engineered target vectors maximizes the use of the *de novo* generated GT platform  
10 *ACes* for most gene targets. Furthermore, using lambda integrase technology, GT platform *ACes* containing multiple *attP* sites permits the opportunity to incorporate multiple therapeutic targets onto a single platform. This could be of value in cases where a defined therapy requires multiple gene targets, a single therapeutic target requires an  
15 additional gene regulatory factor or a GT *ACes* requires a "kill" switch.

Similar to the CPP *ACes*, a feature of the gene target vector is the *promoterless* marker gene containing an upstream *attB* site (marker 2 on Figure 9). Normally, the marker (in this case, a cell surface antigen that can be sorted by FACS would be ideal) would not be expressed unless it  
20 is placed downstream of a promoter sequence. Using the lambda integrase technology ( $\lambda$ INT<sub>E174R</sub> on figure 9), site-specific recombination between the *attB* site on the vector and the promoter- *attP* site on the GT platform *ACes* results in the expression of marker#2 on the gene target vector, i.e. positive selection for the site-specific event. Site-specific  
25 recombination events on the GT *ACes* versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR.

For expression of the therapeutic gene, human specific promoters, such as a ferritin heavy chain promoter (SEQ ID NO:128); EF1 $\alpha$  or RNA

-75-

Poll, are used. These promoters are for high level expression of a cDNA encoded therapeutic protein. In addition to expressing cDNA (or even hybrid cDNA/artificial intron constructs), the GT platform ACes are used for engineering and expressing large genomic fragments carrying  
5 therapeutic genes of interest expressed from native promoter sequences. This is of importance in situations where the therapy requires precise cell specific expression or in instances where expression is best achieved from genomic clones versus cDNA.

10                   **3.     Selectable markers for use, for example, in Gene Therapy (GT)**

The following are selectable markers that can be incorporated into human ACes and used for selection.

15                   **Dual Resistance to 4-Hydroperoxycyclophosphamide and Methotrexate by Retroviral Transfer of the Human Aldehyde Dehydrogenase Class 1 Gene and a Mutated Dihydrofolate Reductase Gene**

The genetic transfer of drug resistance to hematopoietic cells is one approach to overcoming myelosuppression caused by high-dose chemotherapy. Because cyclophosphamide (CTX) and methotrexate  
20 (MTX) are commonly used non-cross-resistant drugs, generation of dual drug resistance in hematopoietic cells that allows dose intensification may increase anti-tumor effects and circumvent the emergence of drug-resistant tumors, a retroviral vector containing a human cytosolic ALDH-1-encoding DNA clone and a human doubly mutated DHFR-encoding  
25 clone (Phe22/Ser31; termed F/S in the description of constructs) to generate increased resistance to CTX and MTX were constructed (Takebe *et al.* (2001) *Mol Ther* 3(1):88-96). This construct may be useful for protecting patients from high-dose CTX- and MTX-induced myelosuppression. ACes can be similarly constructed.

-76-

5                   **Multiple mechanisms of N-phosphonacetyl-L-aspartate resistance in human cell lines: carbamyl-P synthetase/aspartate transcarbamylase/dihydro-orotase gene amplification is frequent only when chromosome 2 is rearranged**

Rodent cells resistant to N-phosphonacetyl-L-aspartate (PALA) invariably contain amplified carbamyl-P synthetase/aspartate transcarbamylase/dihydro-orotase (CAD) genes, usually in widely spaced tandem arrays present as extensions of the same chromosome arm that carries a single copy of CAD in normal cells (Smith *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:1816-21). In contrast, amplification of CAD is very infrequent in several human tumor cell lines. Cell lines with minimal chromosomal rearrangement and with unrearranged copies of chromosome 2 rarely develop intrachromosomal amplifications of CAD. These cells frequently become resistant to PALA through a mechanism that increases the aspartate transcarbamylase activity with no increase in CAD copy number, or they obtain one extra copy of CAD by forming an isochromosome 2p or by retaining an extra copy of chromosome 2. In cells with multiple chromosomal aberrations and rearranged copies of chromosome 2, amplification of CAD as tandem arrays from rearranged chromosomes is the most frequent mechanism of PALA resistance. All of these different mechanisms of PALA resistance are blocked in normal human fibroblasts. Thus, ACes with multiple copies of the CAD gene would provide PALA resistance.

25                   **Retroviral coexpression of thymidylate synthase and dihydrofolate reductase confers fluoropyrimidine and antifolate resistance**

Retroviral gene transfer of dominant selectable markers into hematopoietic cells can be used to select genetically modified cells in vivo or to attenuate the toxic effects of chemotherapeutic agents. Fantz *et al.* ((1998) *Biochem Biophys Res Comm* 243(1):6-12) have shown that

-77-

retroviral gene transfer of thymidylate synthase (TS) confers resistance to TS directed anticancer agents and that co-expression of TS and dihydrofolate reductase (DHFR) confers resistance to TS and DHFR cytotoxic agents. Retroviral vectors encoding *Escherichia coli* TS, human TS, and the Tyr-to-His at residue 33 variant of human TS (Y33HhTS) were constructed and fibroblasts transfected with these vectors conferred comparable resistance to the TS-directed agent fluorodeoxyuridine (FdUrd, approximately 4-fold). Retroviral vectors that encode dual expression of Y33HhTS and the human L22Y DHFR (L22YhDHFR) variants conferred resistance to FdUrd (3- to 5-fold) and trimetrexate (30- to 140-fold). A L22YhDHFR-Y33HhTS chimeric retroviral vector was also constructed and transduced cells were resistant to FdUrd (3-fold), AG337 (3-fold), trimetrexate (100-fold) and methotrexate (5-fold). These results show that recombinant retroviruses can be used to transfer the cDNA that encodes TS and DHFR and dual expression in transduced cells is sufficiently high to confer resistance to TS and DHFR directed anticancer agents. ACes can be similarly constructed.

**Human CD34+ cells do not express glutathione S-transferases alpha**

The expression of glutathione S-transferases alpha (GST alpha) in human hematopoietic CD34+ cells and bone marrow was studied using RT-PCR and immunoblotting (Czerwinski M, Kiem *et al.* (1997) *Gene Ther* 4(3):268-70). The GSTA1 protein conjugates glutathione to the stem cell selective alkylator busulfan. This reaction is the major pathway of elimination of the compound from the human body. Human hematopoietic CD34+ cells and bone marrow do not express GSTA1 message, which was present at a high level in liver, an organ relatively resistant to busulfan toxicity in comparison to bone marrow. Similarly, baboon CD34+ cells and dog bone marrow do not express GSTA1. Thus, human

-78-

GSTA1 is a chemoprotective selectable marker in human stem cell gene therapy and could be employed in *ACes* construction.

**Selection of retrovirally transduced hematopoietic cells using CD24 as a marker of gene transfer**

5           Pawliuk *et al.* ((1994) *Blood* 84(9):2868-2877) have investigated the use of a cell surface antigen as a dominant selectable marker to facilitate the detection and selection of retrovirally infected target cells. The small coding region of the human cell surface antigen CD24 (approximately 240 bp) was introduced into a myeloproliferative sarcoma virus (MPSV)-based retroviral vector, which was then used to infect day 4  
10 5-fluorouracil (5-FU)-treated murine bone marrow cells. Within 48 hours of termination of the infection procedure CD24-expressing cells were selected by fluorescent-activated cell sorting (FACS) with an antibody directed against the CD24 antigen. Functional analysis of these cells  
15 showed that they included not only *in vitro* clonogenic progenitors and day 12 colony-forming unit-spleen but also cells capable of competitive long-term hematopoietic repopulation. Double-antibody labeling studies performed on recipients of retrovirally transduced marrow cells showed that some granulocytes, macrophages, erythrocytes, and, to a lesser  
20 extent, B and T lymphocytes still expressed the transduced CD24 gene at high levels 4 months later. No gross abnormalities in hematopoiesis were detected in mice repopulated with CD24-expressing cells. These results show that the use of the CD24 cell surface antigen as a retrovirally encoded marker permits rapid, efficient, and nontoxic selection *in vitro* of  
25 infected primary cells, facilitates tracking and phenotyping of their progeny, and provides a tool to identify elements that regulate the expression of transduced genes in the most primitive hematopoietic cells. *ACes* could be similarly constructed.

-79-

**DeltahGHR, a biosafe cell surface-labeling molecule for analysis and selection of genetically transduced human cells**

A selectable marker for retroviral transduction and selection of human and murine cells is known (see, Garcia-Ortiz *et al.* (2000) *Hum Gene Ther* 11(2):333-46). The molecule expressed on the cell surface of the transduced population is a truncated version of human growth hormone receptor (deltahGHR), capable of ligand (hGH) binding, but devoid of the domains involved in signal triggering. The engineered molecule is stably expressed in the target cells as an inert protein unable to trigger proliferation or to rescue the cells from apoptosis after ligand binding. This new marker, has a wide application spectrum, since hGHR in the human adult is highly expressed only in liver cells, and lower levels have been reported in certain lymphocyte cell populations. The deltahGHR label has high biosafety potential, as it belongs to a well-characterized hormonal system that is nonessential in adults, and there is extensive clinical experience with hGH administration in humans. The differential binding properties of several monoclonal antibodies (MAbs) are used in a cell rescue method in which the antibody used to select deltahGHR-transduced cells is eluted by competition with hGH or, alternatively biotinylated hGH is used to capture tagged cells. In the latter system, the final purified population is recovered free of attached antibodies in hGH (a substance approved for human use)-containing medium. Such a system could be used to identify ACes containing cells.

**4. Transgenic models for evaluation of genes and discovery of new traits in plants**

Of interest is the use of plants and plant cells containing artificial chromosomes for the evaluation of new genetic combinations and discovery of new traits. Artificial chromosomes, by virtue of the fact that they can contain significant amounts of DNA can also therefore encode

-80-

numerous genes and accordingly a multiplicity of traits. It is contemplated here that artificial chromosomes, when formed from one plant species, can be evaluated in a second plant species. The resultant phenotypic changes observed, for example, can indicate the nature of the genes contained within the DNA contained within the artificial chromosome, and hence permit the identification of novel genetic activities. Artificial chromosomes containing euchromatic DNA or partially containing euchromatic DNA can serve as a valuable source of new traits when transferred to an alien plant cell environment. For example, it is contemplated that artificial chromosomes derived from dicot plant species can be introduced into monocot plant species by transferring a dicot artificial chromosome. The dicot artificial chromosome possessing a region of euchromatic DNA containing expressed genes.

The artificial chromosomes can be designed to allow the artificial chromosome to recombine with the naturally occurring plant DNA in such a fashion that a large region of naturally occurring plant DNA becomes incorporated into the artificial chromosome. This allows the artificial chromosome to contain new genetic activities and hence carry novel traits. For example, an artificial chromosome can be introduced into a wild relative of a crop plant under conditions whereby a portion of the DNA present in the chromosomes of the wild relative is transferred to the artificial chromosome. After isolation of the artificial chromosome, this naturally occurring region of DNA from the wild relative, now located on the artificial chromosome can be introduced into the domesticated crop species and the genes encoded within the transferred DNA expressed and evaluated for utility. New traits and gene systems can be discovered in this fashion. The artificial chromosome can be modified to contain sequences that promote homologous recombination within plant cells, or

-81-

be modified to contain a genetic system that functions as a site-specific recombination system.

Artificial chromosomes modified to recombine with plant DNA offer many advantages for the discovery and evaluation of traits in different  
5 plant species. When the artificial chromosome containing DNA from one plant species is introduced into a new plant species, new traits and genes can be introduced. This use of an artificial chromosome allows for the ability to overcome the sexual barrier that prevents transfer of genes from one plant species to another species. Using artificial chromosomes in this  
10 fashion allows for many potentially valuable traits to be identified including traits that are typically found in wild species. Other valuable applications for artificial chromosomes include the ability to transfer large regions of DNA from one plant species to another, such as DNA encoding potentially valuable traits such as altered oil, carbohydrate or protein  
15 composition, multiple genes encoding enzymes capable of producing valuable plant secondary metabolites, genetic systems encoding valuable agronomic traits such as disease and insect resistance, genes encoding functions that allow association with soil bacterium such as growth promoting bacteria or nitrogen fixing bacteria, or genes encoding traits  
20 that confer freezing, drought or other stress tolerances. In this fashion, artificial chromosomes can be used to discover regions of plant DNA that encode valuable traits.

The artificial chromosome can also be designed to allow the transfer and subsequent incorporation of these valuable traits now located  
25 on the artificial chromosome into the natural chromosomes of a plant species. In this fashion the artificial chromosomes can be used to transfer large regions of DNA encoding traits normally found in one plant species into another plant species. In this fashion, it is possible to derive a plant cell that no longer needs to carry an artificial chromosome to



-82-

posses the novel trait. Thus, the artificial chromosome would serve as the transfer mechanism to permit the formation of plants with greater degree of genetic diversity.

The design of an artificial chromosome to accomplish the afore-  
5 mentioned purposes can include within the artificial chromosome the presence of specific DNA sequences capable of acting as sites for homologous recombination to take place. For example, the DNA sequence of *Arabidopsis* is now known. To construct an artificial chromosome capable of recombining with a specific region of *Arabidopsis*  
10 DNA, a sequence of *Arabidopsis* DNA, normally located near a chromosomal location encoding genes of potential interest can be introduced into an artificial chromosome by methods provided herein. It may be desirable to include a second region of DNA within the artificial chromosome that provides a second flanking sequence to the region  
15 encoding genes of potential interest, to promote a double recombination event which would ensure transfer of the entire chromosomal region, encoding genes of potential interest, to the artificial chromosome. The modified artificial chromosome, containing the DNA sequences capable of homologous recombination region, can then be introduced into  
20 *Arabidopsis* cells and the homologous recombination event selected.

It is convenient to include a marker gene to allow for the selection of a homologous recombination event. The marker gene is preferably inactive unless activated by an appropriate homologous recombination event. For example, US 5,272,071, describes a method where an  
25 inactive plant gene is activated by a recombination event such that desired homologous recombination events can be easily scored. Similarly, US 5,501,967 describes a method for the selection of homologous recombination events by activation of a silent selection gene first introduced into the plant DNA, the gene being activated by an appropriate

-83-

homologous recombination event. Both of these methods can be applied to enable a selective process to be included to select for recombination between an artificial chromosome and a plant chromosome. Once the homologous recombination event is detected, the artificial chromosome, once selected, is isolated and introduced into a recipient cell, for example, tobacco, corn, wheat or rice, and the expression of the newly introduced DNA sequences evaluated.

Phenotypic changes in the recipient plant cells containing the artificial chromosome, or in regenerated plants containing the artificial chromosome, allows for the evaluation of the nature of the traits encoded by the *Arabidopsis* DNA, under conditions naturally found in plant cells, including the naturally occurring arrangement of DNA sequences responsible for the developmental control of the traits in the normal chromosomal environment.

Traits such as durable fungal or bacterial disease resistance, new oil and carbohydrate compositions, valuable secondary metabolites such as phytosterols, flavonoids, efficient nitrogen fixation or mineral utilization, resistance to extremes of drought, heat or cold are all found within different populations of plant species and are often governed by multiple genes. The use of single gene transformation technologies does not permit the evaluation of the multiplicity of genes controlling many valuable traits. Thus, incorporation of these genes into artificial chromosomes allows the rapid evaluation of the utility of these genetic combinations in heterologous plant species.

The large scale order and structure of the artificial chromosome provides a number of unique advantages in screening for new utilities or novel phenotypes within heterologous plant species. The size of new DNA that can be carried by an artificial chromosome can be millions of base pairs of DNA, representing potentially numerous genes that may

-84-

have novel utility in a heterologous plant cell. The artificial chromosome is a "natural" environment for gene expression, the problems of variable gene expression and silencing seen for genes transferred by random insertion into a genome should not be observed. Similarly, there is no  
5 need to engineer the genes for expression, and the genes inserted would not need to be recombinant genes. Thus, one expects the expression from the transferred genes to be temporal and spatial, as observed in the species from where the genes were initially isolated. A valuable feature for these utilities is the ability to isolate the artificial chromosomes and to  
10 further isolate, manipulate and introduce into other cells artificial chromosomes carrying unique genetic compositions.

Thus, the use of artificial chromosomes and homologous recombination in plant cells can be used to isolate and identify many valuable crop traits.

15 In addition to the use of artificial chromosomes for the isolation and testing of large regions of naturally occurring DNA, methods for the use of artificial chromosomes and cloned DNA are also contemplated. Similar to that described above, artificial chromosomes can be used to carry large regions of cloned DNA, including that derived from other plant species.

20 The ability to incorporate novel DNA elements into an artificial chromosome as it is being formed allows for the development of artificial chromosomes specifically engineered as a platform for testing of new genetic combinations, or "genomic" discoveries for model species such as *Arabidopsis*. It is known that specific "recombinase" systems can be  
25 used in plant cells to excise or re-arrange genes. These same systems can be used to derive new gene combinations contained on an artificial chromosome.

The artificial chromosomes can be engineered as platforms to accept large regions of cloned DNA, such as that contained in Bacterial

-85-

Artificial Chromosomes (BACs) or Yeast Artificial Chromosomes (YACs). It is further contemplated, that as a result of the typical structure of artificial chromosomes containing tandemly repeated DNA blocks, that sequences other than cloned DNA sequence can be introduced by recombination processes. In particular recombination within a predefined region of the tandemly repeated DNA within the artificial chromosome provides a mechanism to "stack" numerous regions of cloned DNA, including large regions of DNA contained within BACs or YACs clones. Thus, multiple combinations of genes can be introduced onto artificial chromosomes and these combinations tested for functionality. In particular, it is contemplated that multiple YACs or BACs can be stacked onto an artificial chromosomes, the BACs or YACs containing multiple genes of complex pathways or multiple genetic pathways. The BACs or YACs are typically selected based on genetic information available within the public domain, for example from the Arabidopsis Information Management System (<http://aims.cps.msu.edu/aims/index.html>) or the information related to the plant DNA sequences available from the Institute for Genomic Research (<http://www.tigr.org>) and other sites known to those skilled in the art. Alternatively, clones can be chosen at random and evaluated for functionality. It is contemplated that combinations providing a desired phenotype can be identified by isolation of the artificial chromosome containing the combination and analyzing the nature of the inserted cloned DNA.

In this regard, it is contemplated that the use of site-specific recombination sequences can have considerable utility in developing artificial chromosomes containing DNA sequences recognized by recombinase enzymes and capable of accepting DNA sequences containing same. The use of site-specific recombination as a means to target an introduced DNA to a specific locus has been demonstrated in

-86-

the art and such methods can be employed. The recombinase systems can also be used to transfer the cloned DNA regions contained within the artificial chromosome to the naturally occurring plant or mammalian chromosomes.

5           As noted herein, many site-specific recombinases are known and can be identified (Kilby *et al.* (1993) *Trends in Genetics* 9:413-418). The three recombinase systems that have been extensively employed include: an activity identified as R encoded by the pSR1 plasmid of *Zygosaccharomyces rouxii*, FLP encoded for the 2 $\mu$ m circular plasmid from  
10 *Saccharomyces cerevisiae* and *Cre-lox* from the phage P1.

          The integration function of site-specific recombinases is contemplated as a means to assist in the derivation of genetic combinations on artificial chromosomes. In order to accomplish this, it is contemplated that a first step of introducing site-specific recombinase  
15 sites into the genome of a plant cell in an essentially random manner is conducted, such that the plant cell has one or more site-specific recombinase recognition sequences on one or more of the plant chromosomes. An artificial chromosome is then introduced into the plant cell, the artificial chromosome engineered to contain a recombinase  
20 recognition site (e.g., integration site) capable of being recognized by a site-specific recombinase. Optionally, a gene encoding a recombinase enzyme is also included, preferably under the control of an inducible promoter. Expression of the site-specific recombinase enzyme in the plant cell, either by induction of a inducible recombinase gene, or  
25 transient expression of a recombinase sequence, causes a site-specific recombination event to take place, leading to the insertion of a region of the plant chromosomal DNA (containing the recombinase recognition site) into the recombinase recognition site of the artificial chromosome, and forming an artificial chromosome containing plant chromosomal DNA.

-87-

The artificial chromosome can be isolated and introduced into a heterologous host, preferably a plant host, and expression of the newly introduced plant chromosomal DNA can be monitored and evaluated for desirable phenotypic changes. Accordingly, carrying out this

5 recombination with a population of plant cells wherein the chromosomally located recombinase recognition site is randomly scattered throughout the chromosomes of the plant, can lead to the formation of a population of artificial chromosomes, each with a different region of plant chromosomal DNA, and each potentially representing a novel genetic combination.

10 This method requires the precise site-specific insertion of chromosomal DNA into the artificial chromosome. This precision has been demonstrated in the art. For example, Fukushige and Sauer ((1992) Proc. Natl. Acad. Sci. USA, 89:7905-7909) demonstrated that the *Cre-lox* homologous recombination system could be successfully employed to  
15 introduce DNA into a predefined locus in a chromosome of mammalian cells. In this demonstration a promoter-less antibiotic resistance gene modified to include a *lox* sequence at the 5' end of the coding region was introduced into CHO cells. Cells were re-transformed by electroporation with a plasmid that contained a promoter with a *lox* sequence and a  
20 transiently expressed *Cre* recombinase gene. Under the conditions employed, the expression of the *Cre* enzyme catalyzed the homologous recombination between the *lox* site in the chromosomally located promoter-less antibiotic resistance gene, and the *lox* site in the introduced promoter sequence, leading to the formation of a functional antibiotic  
25 resistance gene. The authors demonstrated efficient and correct targeting of the introduced sequence, 54 of 56 lines analyzed corresponded to the predicted single copy insertion of the DNA due to *Cre* catalyzed site-specific homologous recombination between the *lox* sequences.

-88-

Accordingly a *lox* sequence may be first added to a genome of a plant species capable of being transformed and regenerated to a whole plant to serve as a recombinase target DNA sequence for recombination with an artificial chromosome. The *lox* sequence may be optimally  
5 modified to further contain a selectable marker which is inactive but can be activated by insertion of the *lox* recombinase recognition sequence into the artificial chromosome.

A promoterless marker gene or selectable marker gene linked to the recombinase recognition sequence, which is first inserted into the  
10 chromosomes of a plant cell can be used to engineer a platform chromosome. A promoter is linked to a recombinase recognition site, in an orientation that allows the promoter to control the expression of the marker or selectable marker gene upon recombination within the artificial chromosome. Upon a site-specific recombination event between a  
15 recombinase recognition site in a plant chromosome and the recombinase recognition site within the introduced artificial chromosome, a cell is derived with a recombined artificial chromosome, the artificial chromosome containing an active marker or selectable marker activity that permits the identification and or selection of the cell.

20 The artificial chromosomes can be transferred to other plant or animal species and the functionality of the new combinations tested. The ability to conduct such an inter-chromosomal transfer of sequences has been demonstrated in the art. For example, the use of the *Cre-lox* recombinase system to cause a chromosome recombination event  
25 between two chromatids of different chromosomes has been shown.

Any number of recombination systems may be employed as described herein, such as, but not limited to, bacterially derived systems such as the att/int system of phage lambda, and the Gin/gix system.

-89-

More than one recombination system may be employed, including, for example, one recombinase system for the introduction of DNA into an artificial chromosome, and a second recombinase system for the subsequent transfer of the newly introduced DNA contained within an  
5 artificial chromosome into the naturally occurring chromosome of a second plant species. The choice of the specific recombination system used will be dependent on the nature of the modification contemplated.

By having the ability to isolate an artificial chromosome, in particular, artificial chromosomes containing plant chromosomal DNA  
10 introduced via site-specific recombination, and re-introduce the chromosome into other mammalian or plant cells, particularly plant cells, these new combinations can be evaluated in different crop species without the need to first isolate and modify the genes, or carry out multiple transformations or gene transfers to achieve the same  
15 combination isolation and testing combinations of the genes in plants. The use of a site-specific recombinase also allows the convenient recovery of the plant chromosomal region into other recombinant DNA vectors and systems, such as mammalian or insect systems, for manipulation and study.

20 Also contemplated herein are *ACes*, cell lines and methods for use in screening a new chromosomal combinations, deletions, truncations with eucaryotic genome that take advantage of the site-specific recombination systems incorporated onto platform *ACes* provided herein. For example, provided herein is a cell line useful for making a library of  
25 *ACes*, comprising a multiplicity of heterologous recombination sites randomly integrated throughout the endogenous chromosomes. Also provided herein is a method of making a library of *ACes* comprising random portions of a genome, comprising introducing one or more *ACes* into a cell line comprising a multiplicity of heterologous recombination



-90-

- sites randomly integrated throughout the endogenous chromosomes, under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites within the cell's chromosomal DNA; and isolating said
- 5 multiplicity of *ACes*, thereby producing a library of *ACes* whereby multiple *ACes* have different portions of the genome within. Also provided herein is a library of cells useful for genomic screening, said library comprising a multiplicity of cells, wherein each cell comprises an *ACes* having a mutually exclusive portion of a chromosomal nucleic acid therein. The
- 10 library of cells can be from a different species and/or cell type than the chromosomal nucleic acid within the *ACes*. Also provided is a method of making one or more cell lines, comprising
- a) integrating into endogenous chromosomal DNA of a selected cell species, a multiplicity of heterologous recombination sites,
  - 15 b) introducing a multiplicity of *ACes* under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites integrated within the cell's endogenous chromosomal DNA;
  - c) isolating said multiplicity of *ACes*, thereby producing a library of
  - 20 *ACes* whereby a multiplicity of *ACes* have mutually exclusive portions of the endogenous chromosomal DNA therein;
  - d) introducing the isolated multiplicity of *ACes* of step c) into a multiplicity of cells, thereby creating a library of cells;
  - e) selecting different cells having mutually exclusive *ACes* therein
  - 25 and clonally expanding or differentiating said different cells into clonal cell cultures, thereby creating one or more cell lines.

These *ACes*, cell lines and methods utilize the site-specific recombination sites on platform *ACes* analogous YAC manipulation related to: the methods of generating terminal deletions in normal and

-91-

artificial chromosomes (e.g., *ACes*; as described in Vollrath et al., 1988, *PNAS, USA*, 85:6027-66031; and Pavan et al., *PNAS, USA*, 87:1300-1304 ); the methods of generating interstitial deletions in normal and artificial chromosomes (as described in Campbell et al., 1991, *PNAS, USA*, 888:5744-5748); and the methods of detecting homologous recombination between two *ACes* (as described in Cellini et al., 1991, *Nuc. Acid Res.*, 19(5):997-1000).

**5. Use of platform *ACes* in Pharmacogenomic/toxicology applications (development of "Reporter *ACes*")**

10 In addition to the placement of genes onto *ACes* chromosomes for therapeutic protein production or gene therapy, the platform can be engineered via the *IntR* lambda integrase to carry reporter-linked constructs (reporter genes) that monitor changes in cellular physiology as measured by the particular reporter gene (or a series of different reporter  
15 genes) readout. The reporter linked constructs are designed to include a gene that can be detected (by for example fluorescence, drug resistance, immunohistochemistry, or transcript production, and the like) with well-known regulatory sequences that would control the expression of the detectable gene. Exemplary regulatory promoter sequences are well-  
20 known in the art:

**A) Reporter *ACes* for drug pathway screening**

The *ACes* can be engineered to carry reporter-linked constructs that indicate a signal is being transduced through one or a number of pathways. For example, transcriptionally regulated promoters from genes  
25 at the end (or any other chosen point) of particular signal transduction pathways could be engineered on the *ACes* to express the appropriate readout (either by fluorescent protein production or drug resistance) when the pathway is activated (or down-regulated as well). In one embodiment, a number of reporters from different can be placed on a

-92-

*ACes* chromosome. Cells (and/or whole animals) containing such a Reporter *ACes* could be exposed to a variety of drugs or compounds and monitored for the effects of the drugs or compounds upon the selected pathway(s) by the reporter gene(s). Thus, drugs or compounds can be  
5 classified or identified by particular pathways they excite or down-regulate. Similarly, transcriptional profiles obtained from genomic array experiments can be biologically validated using the reporter *ACes* provided herein.

**B) Reporter *ACes* for toxic compound testing**

10 Environmental or man-made genotoxicants can be tested in cell lines carrying a number of reporter-genes platform *ACes* linked to promoters that are transcriptionally regulated in response to DNA damage, induced apoptosis or necrosis, and cell-cycle perturbations. Furthermore, new drugs and/or compounds could be tested in a similar manner with the  
15 genotoxicant *ACes* reporter for their cellular/genetic toxicity by such a screen. Likewise, toxic compound testing could be carried out in whole transgenic animals carrying the *ACes* chromosome that measures genotoxicant exposure ("canary in a coal mine"). Thus, the same or similar type *ACes* could be used for toxicity testing in either a cell-based  
20 or whole animal setting. An example would include *ACes* that carry reporter-linked genes controlled by various cytochrome P450 profiled promoters and the like.

**C) Reporter *ACes* for individualized pharmacogenomics/drug profiling**

25 A common disease may arise via various mechanisms. In many instances there are multiple treatments available for a given disease. However, the success of a given treatment may depend upon the mechanism by which the disease originated and/or by the genetic background of the patient. In order to establish the most effective

-93-

treatment for a given patient one could utilize the *ACes* reporters provided herein. *ACes* reporters can be used in patient cell samples to determine an individualized drug regimen for the patient. In addition, potential polymorphisms affecting the transcriptional regulation of an individual's particular gene can be assessed by this approach.

**D) Reporter *ACes* for classification of similar patient tumors**

As with other diseases as described in 5.C) above, cancer cells arise via different mechanisms. Furthermore, as a cancerous cell propagates it may undergo genomic alterations. An *ACes* reporter transferred to cells of different patients having the same disease, i.e. similar cancers, could be used to categorize the particular cancer of each patient, thereby facilitating the identification of the most effective therapeutic regimen. Examples would include the validation of array profiling of certain classes of breast cancers. Subsequently, appropriate drug profiling could be carried out as described above.

**E) Reporter *ACes* as a "differentiation" sensor**

Using the *ACes* reporter as a "differentiation" sensor in stem cells or other progenitor cells in order to enrich by selection (either FACS based screening, drug selection and/or use of suicide gene) for a particular class of differentiated or undifferentiated cells. For example, in one embodiment, this assay could also be used for compound screening for small molecule modifiers of cell differentiation.

**F) Whole animal studies with Reporter *ACes***

Finally, with whole-body fluorescence imaging technology (Yang et al. (2000) PNAS 97:12278) any of the above Reporter *ACes* methods could be used in conjunction with whole-body imaging to monitor reporter genes within whole animals without sacrificing the animal. This would allow temporal and spatial analysis of expression patterns under a given set of conditions. The conditions tested may include for example, normal

-94-

differentiation of a stem cell, response to drug or compound treatment whether targeted to the diseased tissue or presented systemically, response to genotoxics, and the like.

The following examples are included for illustrative purposes only  
5 and are not intended to limit the scope of the invention.

#### EXAMPLE 1

##### pFK161

Cosmid pFK161 (SEQ ID NO: 118) was obtained from Dr. Gyula Hadlaczký and contains a 9 kb *NotI* insert derived from a murine rDNA  
10 repeat (see clone 161 described in PCT Application Publication No. WO97/40183 by Hadlaczký *et al.* for a description of this cosmid). This cosmid, referred to as clone 161 contains sequence corresponding to nucleotides 10,232-15,000 in SEQ ID NO. 26. It was produced by inserting fragments of the megachromosome (see, U.S. Patent No.  
15 6,077,697 and International PCT application No. WO 97/40183). For example, H1D3, which was deposited at the European Collection of Animal Cell Culture (ECACC) under Accession No. 96040929, is a mouse-hamster hybrid cell line carrying this megachromosome into plasmid pWE15 (Stratagene, La Jolla, California; SEQ ID No. 31) as  
20 follows. Half of a 100  $\mu$ l low melting point agarose block (mega-plug) containing isolated SATACs was digested with *NotI* overnight at 37°C. Plasmid pWE15 was similarly digested with *NotI* overnight. The mega-plug was then melted and mixed with the digested plasmid, ligation buffer and T4 DNA ligase. Ligation was conducted at 16°C overnight. Bacterial  
25 DH5 $\alpha$  cells were transformed with the ligation product and transformed cells were plated onto LB/Amp plates. Fifteen to twenty colonies were grown on each plate for a total of 189 colonies. Plasmid DNA was isolated from colonies that survived growth on LB/Amp medium and analyzed by Southern blot hybridization for the presence of DNA that

-95-

hybridized to a pUC19 probe. This screening methodology assured that all clones, even clones lacking an insert but yet containing the pWE15 plasmid, would be detected.

Liquid cultures of all 189 transformants were used to generate cosmid minipreps for analysis of restriction sites within the insert DNA. Six of the original 189 cosmid clones contained an insert. These clones were designated as follows: 28 (~9-kb insert), 30 (~9-kb insert), 60 (~4-kb insert), 113 (~9-kb insert), 157 (~9-kb insert) and 161 (~9-kb insert). Restriction enzyme analysis indicated that three of the clones (113, 157 and 161) contained the same insert.

For sequence analysis the insert of cosmid clone no. 161 was subcloned as follows. To obtain the end fragments of the insert of clone no. 161, the clone was digested with *NotI* and *Bam*HI and ligated with *NotI*/*Bam*HI-digested pBluescript KS (Stratagene, La Jolla, California). Two fragments of the insert of clone no. 161 were obtained: a 0.2-kb and a 0.7-kb insert fragment. To subclone the internal fragment of the insert of clone no. 161, the same digest was ligated with *Bam*HI-digested pUC19. Three fragments of the insert of clone no. 161 were obtained: a 0.6-kb, a 1.8-kb and a 4.8-kb insert fragment.

The insert corresponds to an internal section of the mouse ribosomal RNA gene (rDNA) repeat unit between positions 7551-15670 as set forth in GENBANK accession no. X82564, which is provided as SEQ ID NO. 18. The sequence data obtained for the insert of clone no. 161 is set forth in SEQ ID NOS. 19-25. Specifically, the individual subclones corresponded to the following positions in GENBANK accession no. X82564 (SEQ ID NO:18) and in SEQ ID NOS. 19-25:

-96-

5

Subclone	Start	End	Site	SEQ ID No.
	in X82564			
161k1	7579	7755	<i>NotI, BamHI</i>	19
161m5	7756	8494	<i>BamHI</i>	20
161m7	8495	10231	<i>BamHI</i>	21 (shows only sequence corresponding to nt. 8495-8950), 22 (shows only sequence corresponding to nt. 9851-10231)
161m12	10232	15000	<i>BamHI</i>	23 (shows only sequence corresponding to nt. 10232-10600), 24 (shows only sequence corresponding to nt. 14267-15000)
161k2	15001	15676	<i>NotI, BamHI</i>	25

10 The sequence set forth in SEQ ID NOs. 19-25 diverges in some positions from the sequence presented in positions 7551-15670 of GENBANK accession no. X82564. Such divergence may be attributable to random mutations between repeat units of rDNA.

15 For use herein, the rDNA insert from the clone was prepared by digesting the cosmid with *NotI* and *BglII* and was purified as described above. Growth and maintenance of bacterial stocks and purification of plasmids were performed using standard well known methods (see, *e.g.*, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press), and plasmids were purified from bacterial cultures using Midi - and Maxi-preps Kits (Qiagen, 20 Mississauga, Ontario).

#### pDsRed1N1

25 This vector is available from Clontech (see SEQ ID No. 29) and encodes the red fluorescent protein (DsRed; Genbank accession no. AF272711; SEQ ID Nos. 39 and 40). DsRed, which has a vivid red fluorescence, was isolated from the IndoPacific sea anemone relative *Discosoma* species. The plasmid pDsRed1N1 (Clontech; SEQ ID No. 29) constitutively expresses a human codon-optimized variant of the

-97-

fluorescent protein under control of the CMV promoter. Unmodified, this vector expresses high levels of DsRed1 and includes sites for creating N-terminal fusions by cloning proteins of interest into the multiple cloning site (MCS). It is Kan and Neo resistant for selection in bacterial or  
5 eukaryotic cells.

#### Plasmid pMG

Plasmid pMG (InvivoGen, San Diego, California; see SEQ. ID. NO. 27 for the nucleotide sequence of pMG) contains the hygromycin phosphotransferase gene under the control of the immediate-early human  
10 cytomegalovirus (hCMV) enhancer/promoter with intron A. Vector pMG also contains two transcriptional units allowing for the coexpression of two heterologous genes from a single vector sequence.

The first transcriptional unit of pMG contains a multiple cloning site for insertion of a gene of interest, the hygromycin phosphotransferase  
15 gene (*hph*) and the immediate-early human cytomegalovirus (hCMV) enhancer/promoter with intron A (see, *e.g.*, Chapman *et al.* (1991) *Nuc. Acids Res.* 19:3979-3986) located upstream of *hph* and the multiple cloning site, which drives the expression of *hph* and any gene of interest inserted into the multiple cloning site as a polycistronic mRNA. The first  
20 transcriptional unit also contains a modified EMCV internal ribosomal entry site (IRES) upstream of the *hph* gene but downstream of the hCMV promoter and MCS for ribosomal entry in translation of the *hph* gene (see SEQ ID NO. 27, nucleotides 2736-3308). The IRES is modified by insertion of the constitutive *E. coli* promoter (EM7) within an intron (IM7)  
25 into the end of the IRES. In mammalian cells, the *E. coli* promoter is treated as an intron and is spliced out of the transcript. A polyadenylation signal from the bovine growth hormone (bGh) gene (see, *e.g.*, Goodwin and Rottman (1992) *J. Biol. Chem.* 267:16330-16334) and a pause site derived from the 3' flanking region of the human  $\alpha 2$



-98-

globin gene (see, *e.g.*, Enriquez-Harris *et al.* (1991) *EMBO J.* 10:1833-1842) are located at the end of the first transcription unit. Efficient polyadenylation is facilitated by inserting the flanking sequence of the bGh gene 3' to the standard AAUAAA hexanucleotide sequence.

5        The second transcriptional unit of pMG contains another multiple cloning site for insertion of a gene of interest and an EF-1 $\alpha$ /HTLV hybrid promoter located upstream of this multiple cloning site, which drives the expression of any gene of interest inserted into the multiple cloning site. The hybrid promoter is a modified human elongation factor-1 alpha (EF-1  
10    alpha) gene promoter (see, *e.g.*, Kim *et al.* (1990) *Gene* 91:217-223) that includes the R segment and part of the U5 sequence (R-U5') of the human T-cell leukemia virus (HTLV) type I long terminal repeat (see, *e.g.*, Takebe *et al.* (1988) *Mol. Cell. Biol.* 8:466-472). The Simian Virus 40 (SV40) late polyadenylation signal (see Carswell and Alwine (1989) *Mol.*  
15    *Cell. Biol.* 9:4248-4258) is located downstream of the multiple cloning site. Vector pMG contains a synthetic polyadenylation site for the first and second transcriptional units at the end of the transcriptional unit based on the rabbit  $\beta$ -globin gene and containing the AATAAA hexanucleotide sequence and a GT/T-rich sequence with 22-23  
20    nucleotides between them (see, *e.g.*, Levitt *et al.* (1989) *Genes Dev.* 3:1019-1025). A pause site derived from the C2 complement gene (see, Moreira *et al.* (1995) *EMBO J.* 14:3809-3819) is also located at the 3' end of the second transcriptional unit.

Vector pMG also contains an ori sequence (ori pMB1) located  
25    between the SV40 polyadenylation signal and the synthetic polyadenylation site.

## EXAMPLE 2

**A. Construction of targeting vector and transfection into LMtk- cells for the generation of platform chromosomes**

-99-

- A targeting vector derived from the vector pWE15 (GeneBank Accession # X65279) was modified by replacing the *Sa/I* (Klenow filled)/*Sma*I neomycin resistance containing fragment with the *Pvu*II/*Bam*HI (Klenow filled) puromycin resistance containing fragment
- 5 (isolated from plasmid pPUR, Clontech Laboratories, Inc. Palo Alto, CA; SEQ ID No. 30) resulting in plasmid pWEPuro. Subsequently a 9 Kb *Not*I fragment from the plasmid pFK161 (SEQ ID NO: 118) containing a portion of the mouse rDNA region was cloned into the *Not*I site of pWEPuro resulting in plasmid pWEPuro9K (Figure 2). The vector pWEPuro9K was
- 10 digested with *Spe*I to linearize and transfected into LMtk- mouse cells. Puromycin resistant colonies were isolated and subsequently tested for artificial chromosome formation via fluorescent *in situ* hybridization (FISH) (using mouse major and minor DNA repeat sequences, the puromycin gene and telomeres sequences as probes), and fluorescent activated cell
- 15 sorting (FACS). From this sort, a subclone was isolated containing an artificial chromosome, designated 5B11.12, which carries 4-8 copies of the puromycin resistance gene contained on the pWEPuro9K vector. FISH analysis of the 5B11.12 subclone demonstrated the presence of telomeres and mouse minor on the ACes. DOT PCR has been done on
- 20 the 5B11.12 ACes revealing the absence of uncharacterized euchromatic regions on the ACes. A recombination site, such as an *att* or *loxP* engineering site or a plurality thereof, was introduced onto this ACes thereby providing a platform for site-specific introduction of heterologous nucleic acid.
- 25 **B. Targeting a single sequence specific recombination site onto platform chromosomes**

After the generation of the 5B11.12 platform, a single sequence-specific recombination site is placed onto the platform chromosome via homologous recombination. For this, DNA sequences containing the site-

-100-

specific recombination sequence can be flanked with DNA sequences of homology to the platform chromosome. For example, using the platform chromosome made from the pWEPuro9K vector, mouse rDNA sequences or mouse major satellite DNA can be used as homologous sequences to

5 target onto the platform chromosome. A vector is designed to have these homologous sequences flanking the site-specific recombination site and, after the appropriate restriction enzyme digest to generate free ends of homology to the platform chromosome, the DNA is transfected into cells harboring the platform chromosome (Figure 3). Examples of site-specific

10 cassettes that are targeted to the platform chromosome using either mouse rDNA or mouse major repeat DNA include the SV40-attP-hygro cassette and a red fluorescent protein (RFP) gene flanked by loxP sites (Cre/lox, see, *e.g.*, U.S. Patent No. 4,959,317 and description herein). After transfection and integration of the site-specific cassette,

15 homologous recombination events onto the platform chromosome are subcloned and identified by FACS (*e.g.* screen and single cell subclone via expression of resistance or fluorescent marker) and PCR analysis.

For example, a vector can be constructed containing regions of the mouse rDNA locus flanking a gene cassette containing the SV40 early

20 reporter-bacteriophage lambda attP site-hygromycin selectable marker (see Figure 4 and described below). The use of the bacteriophage lambda attP site for lambda integrase-mediated site-specific recombination is described below. Homologous recombination event of the SV40-attP-hygro cassette onto the platform chromosome was identified using PCR

25 primers that detect the homologous recombination and further confirmed by FISH analysis. After identifying subcloned colonies containing the platform chromosome with a single site-specific recombination site, cells carrying the platform chromosome with a single site-specific

-101-

recombination site can now be engineered with site-specific recombinases (e.g. lambda INT, Cre) for integrating a target gene expression vector.

**C. Targeting a red fluorescent protein (RFP) gene flanked by loxP sites  
5 onto 5B11.12 platform**

As another example, while loxP recombination sites could have been introduced onto the *ACes* during *de novo* biosynthesis, it was thought that this might result in multiple segments of the *ACes* containing a high number of loxP sites, potentially leading to instability upon *Cre*-  
10 mediated recombination. A gene targeting approach was therefore devised to introduce a more limited number of loxP recombination sites into a locus of the 5B11-12 *ACes* containing introduced and possibly co-amplified endogenous rDNA sequences. Although there are more than 200 copies of rDNA genes in the haploid mouse genome distributed  
15 amongst 5-11 chromosomes (depending on strain), rDNA sequences were chosen as the target on the *ACes* since they represent a less frequent target than that of the satellite repeat sequences. Moreover, having observed much stronger pWEPuro9K hybridization to the 5B11-12 *ACes* than to other LMTK<sup>-</sup> chromosomes and in light of the observation that the  
20 transcribed spacer sequences within the rDNA may be less conserved than the rRNA coding regions, it was contemplated that a targeting vector based on the rDNA gene segment in pWEPuro9K would have a higher probability of targeting to the *ACes* rather than to other LMTK<sup>-</sup> chromosomes. Accordingly, a targeting vector, pBSFKLoxDsRedLox, was  
25 designed and constructed based on the rDNA sequences contained in pWEPuro9K.

The plasmid pBSFKLoxDsRedLox was generated in 4 steps. First, the *NotI* rDNA insert of pWEPuro9K (Figure 2) was inserted into pBS SK- (Stratagene) giving rise to pBSFK. Second, a loxP polylinker cassette was

-102-

generated by PCR amplification of pNEB193 (SEQ ID NO:32; New England Biolabs) using primers complementary to the M13 forward and reverse priming sites at their 3' end and a 34 bp 5' extension comprising a LoxP site. This cassette was reinserted into pNEB193 generating  
5 p193LoxMCSLox. Third, the DsRed gene from pDsRed1-N1 (SEQ ID NO:29; Clontech) was then cloned into the polylinker between the loxP sites generating p193LoxDsRedLox. Fourth, a fragment consisting of the DsRed gene flanked by loxP sites was cloned into a unique *NdeI* within the rDNA insert of pBSFK generating pBSFKLoxDsRedLox.

10 A gel purified 11 Kb *PmlI* / *EcoRV* fragment of pBSFKLoxDsRedLox was used for transfection. To detect targeted integration, PCR primers were designed from rDNA sequences within the 5' *NotI*-*PmlI* fragment of pWEPuro9K that is not present on the targeting fragment (5' primer) and sequence within the LoxDsRedLox cassette (3' primer). If the targeting  
15 DNA integrated correctly within the rDNA sequences, PCR amplification using these primers would give rise to a 2.3 Kb band. PCR reactions containing 1-4  $\mu$ l of genomic DNA were carried out according to the MasterTaq protocol (Eppendorf), using murine rDNA 5' primer (5'-CGGACAATGCGGTTGTGCGT-3'; SEQ ID NO:72) and DsRed 3' primer  
20 (5'GGCCCCGTAATGCAGAAGAA-3'; SEQ ID NO:73) and PCR products were analyzed by agarose gel electrophoresis.

1.5X10<sup>6</sup> 5B11-12 LMTK<sup>-</sup> cells were transfected with 2  $\mu$ g of the pBSFKLoxDsRedLox targeting DNA described above using Lipofectamine Plus (Invitrogen). For flow sorting, harvested cells were suspended in  
25 medium and applied to the Becton Dickinson Vantage SE cell sorter, equipped with 488 nm lasers for excitation and 585/42 bandpass filter for optimum detection of RFP fluorescence. Cells were sorted using dPBS as sheath buffer. Negative control parental 5B11-12 cells and a positive control LMTK<sup>-</sup> cell line stably transfected with DsRed were used to

-103-

establish the selection gates. The RFP positive gated populations were recovered, diluted in medium supplemented with 1X penicillin-streptomycin (Invitrogen), then plated and cultured as previously described. After 4 rounds of enrichment, the percentage of RFP positive cells reached levels of 50% or higher. DNA from populations was analyzed by PCR for evidence of targeted integration. Ultimately, single cell subclones were established from positive pools and were analyzed by PCR and PCR-positive clones confirmed by FISH as described below. DNA was purified from pools or single cell clones using previously described methods set forth in Lahm et al., Transgenic Res., 1998; 7:131-134, or in some cases using a Wizard Genomic DNA purification kit (Promega). For FISH analysis, a biotinylated DsRed gene probe was generated by PCR using DsRed specific primers and biotin-labeled dUTP (5' RFP primer: 5'-GGTTTAAAGTGCGCTCCTCCAAGAACGTCATC-3', SEQ ID NO:74; and 3' RFP primer: 5'AGATCTAGAGCCGCGCTACAGGAACAGGTGGTGGCGGCC-3'; SEQ ID NO:75). To maximize the signal intensity of the DsRed probe, Tyramide amplification was carried out according to the manufacturers protocols (NEN).

The process of testing the feasibility of a more general targeting strategy that would not rely on enrichment *via* drug selection of stably transfected clones can be summarized as follows. A red fluorescent protein gene (RFP; encoded by the DsRed gene) was inserted between the loxP sites of the targeting vector to form pBSFKLoxDsRedLox. After transfection with pBSFKLoxDsRedLox, sequential rounds of high speed flow sorting and expansion of sorted cells in culture could then be used to enrich for stable transformants expressing RFP. In the event of targeted integration, PCR screening with primers that amplify from a spacer region within the segment of the 45s pre-rRNA gene in pWEPuro9K to a specific

-104-

anchor sequence within the DsRed gene in the targeting cassette would give rise to a diagnostic 2.3 Kb band. However, as rDNA clusters are found on several chromosomes, confirmation of targeting to an *ACes* would require fluorescence in situ hybridization (FISH) analysis. Finally, the flanking of the DsRed gene by loxP sites would allow for its removal and subsequent replacement with other genes of interest.

After transfection of the targeting sequence into 5B11-12 cells, enrichment for targeted clones was carried out using a combination of flow cytometry to detect red-fluorescing cells and PCR screening.

Ultimately 17 single cell subclones were identified as potential targeted clones by PCR and of these 16 were found by FISH to contain the DsRed integration event into the *ACes*. These subclones are referred to herein as D11-C4, D11-C12, D11-H3, C9-C9, C9-B9, C9-F4, C9-H8, C9-F2, C9-G8, C9-B6, C9-G3, C9-E12, C9-A11, C11-E3, C11-A9 and C11-H4. PCR analysis of genomic DNA isolated from the D11-C4 subclone gave rise to a 2.3 Kb band, indicative of a targeted integration into an rDNA locus. Further analysis of the subclone by FISH analysis with a DsRed gene probe demonstrated integration of the LoxDsRedLox targeting cassette on the *ACes* co-localizing with one of the regions of rDNA staining seen on the 5B11-12 *ACes*, consistent with a targeted integration into an rDNA locus of the *ACes*, while integrations on other chromosomes were not observed. Since transfected cells were maintained as heterogeneous populations through several cycles of sorting and replating it was not possible to estimate the frequency of targeted events. In most mammalian cell lines the frequency of gene targeting via homologous recombination is roughly  $10^{-5}$ - $10^{-7}$  treated cells. Despite the low frequency of these events in mammalian cells, it is clear that an RFP expression based screening paradigm, coupled with PCR analysis, can effectively detect and enrich for such infrequent events in a large

-105-

population. In instances where drug selection is not possible or not desirable, such a system may provide a useful alternative. It was also verified that the modified *ACes* in subclone D11-C4 could be purified by flow cytometry. The results indicate that the flow karyogram of the D11-C4 subclone was unaltered from that of the 5B11-12 cell line. Thus, the D11-C4 *ACes* can be purified in high yield from native chromosomes of the host cell line.

**D. Reduction of LoxP on *ACes* to a single site.**

The strong hybridization signal detected by FISH on the *ACes* using the DsRed gene probe suggests that several copies of the targeting cassette may be present on the *ACes* in the D11-C4 line. This also suggests that multiple rDNA genes have been correctly targeted.

Accordingly, in certain embodiments where necessary, the number of loxP sites on the *ACes* can be reduced to a single site by *in situ* treatment with *Cre* recombinase, provided that the sites are co-linear. Such a process is described for multiple loxP-flanked integrations on a native mouse chromosome (Garrrick et al., Nature Genet., 1998, Jan;18(1):56-59). Reduction to a single loxP site on the D11-C4 *ACes* would result in the loss of the DsRed gene, forming the basis of a useful screen for this event.

For this purpose, a Cre expression plasmid pCX-Cre/GFP III has been generated by first deleting the EcoRI fragment of pCX-eGFP (SEQ ID NO:71) containing the eGFP coding sequence and replacing it with that of a PCR amplified Cre recombinase coding sequence (SEQ ID NO:58), generating pCX-Cre. Next, the Asel/Sspl fragment of pD2eGFP-N1 (containing the CMV promoter driving the D2EGFP gene with SV40 polyA signal; Clontech; SEQ ID NO:87) was inserted into the filled HindIII site of pCX-Cre, generating pCX-Cre\GFP III. Control plasmid pCX-CreRev\GFP



-106-

III was generated in similar fashion except that the Cre recombinase coding sequence was inserted in the antisense orientation. LMTK<sup>-</sup> cell line D11-C4 (containing first generation platform ACes with multiple loxP-DsRED sites) and 5B11-12 cell line (containing ACes with no loxP-DsRED sites) are maintained in culture as described above. D11C4 cells are transfected with 2  $\mu$ g of plasmid pCX-Cre\GFP III or 2  $\mu$ g pCX-CreRev\GFP III using Lipofectamine (Invitrogen) as previously described.

Forty-eight to seventy-two hours after transfection, transfected D11-C4 cells are harvested and GFP positive cells are sorted by cell cytometry using a FACSta Vantage cell sorter (Beckton-Dickinson) as follows: All D11-C4 cells transfected with pCX-Cre\GFP III or control plasmid pCX-CreRev\GFP III that exhibit GFP fluorescent higher than the gate level established by untransfected cells are collected and placed in culture a further 7-14 days. After 7-14 days the initial D11-C4 cells are harvested and analyzed by cell cytometry as follows: Untransfected D11-C4 cells are used to establish the gate that defines the RFP positive population, while 5B11-12 cells are used to set the RFP negative gate. The GFP positive population of D11-C4 transfected with pCX-Cre\GFP III should show decreased red fluorescence compared to pCX-CreRev\GFP III transfected or untransfected control D11-C4 cells. The cells exhibiting greatly decreased or no RFP expression are collected and single cell clones subsequently established. These clones will be expanded and analyzed by fluorescence *in-situ* hybridization and Southern blotting to confirm the removal of loxP-DsRed gene copies.

25

### EXAMPLE 3

**Construction of targeting vector and transfection into LMtk<sup>-</sup> cells for the generation of platform chromosomes containing multiple site-specific recombination sites**

-107-

An example of a selectable marker system for the creation of a chromosome-based platform is shown in Figure 4. This system includes a vector containing the SV40 early promoter immediately followed by (1) a 282 base pair (bp) sequence containing the bacteriophage lambda attP site and (2) the puromycin resistance marker. Initially a *PvuII/StuI* fragment containing the SV40 early promoter from plasmid pPUR (Clontech Laboratories, Inc., Palo Alto, CA; Seq ID No. 30) was subcloned into the *EcoRI/CRI* site of pNEB193 (a PUC19 derivative obtained from New England Biolabs, Beverly, MA; SEQ ID No. 32) generating the plasmid pSV40193. The only differences between pUC19 and pNEB193 are in the polylinker region. A unique *Ascl* site (GGCGCGCC) is located between the *BamHI* site and the *SmaI* site, a unique *PacI* site (TTAATTAA) is located between the *BamHI* site and the *XbaI* site and a unique *PmeI* site (GTTTAAAC) is located between the *PstI* site and the *SaI* site.

The attP site was PCR amplified from lambda genome (GenBank Accession # NC 001416) using the following primers:

attPUP: CCTTGCGCTAATGCTCTGTTACAGG SEQ ID No. 1

attPDWN: CAGAGGCAGGGAGTGGGACAAAATTG SEQ ID No. 2

After amplification and purification of the resulting fragment, the attP site was cloned into the *SmaI* site of pSV40193 and the orientation of the attP site was determined by DNA sequence analysis (plasmid pSV40193attP). The gene encoding puromycin resistance (Puro) was isolated by digesting the plasmid pPUR (Clontech Laboratories, Inc. Palo Alto, CA) with *AgeI/BamHI* followed by filling in the overhangs with Klenow and subsequently cloned into the *Ascl* site downstream of the attP site of pSV40193attP generating the plasmid pSV40193attPsensePUR (Figure 4; SEQ ID NO:113)).

-108-

The plasmid pSV40193attPsensePUR was digested with *ScaI* and co-transfected with the plasmid pFK161 (SEQ ID NO: 118) into mouse LMTk- cells and platform artificial chromosomes were identified and isolated as described above. The process for generating this exemplary platform ACes containing multiple site-specific recombination sites is summarized in Figure 5. One platform ACes resulting from this experiment is designated B19-18. This platform ACes chromosome may subsequently be engineered to contain target gene expression nucleic acids using the lambda integrase mediated site-specific recombination system as described herein in Example 7 and 8.

#### EXAMPLE 4

##### **Lambda integrase mediated site-specific recombination of a RFP expressing vector onto artificial chromosomes**

In this example, a vector expressing the red fluorescent protein (RFP) was produced and recombined into the attP site residing on an artificial chromosome within LMTK- cells. This recombination is depicted in Figure 7.

##### **A. Construction of expression vectors containing wildtype and mutant lambda integrase**

Mutations at the glutamic acid at position 174 in the lambda integrase protein relaxes the requirement for the accessory protein IHF during recombination and DNA supercoiling *in vitro* (see, Miller *et al.* (1980) *Cell* 20:721-729; Lange-Gustafson *et al.* (1984) *J. Biol. Chem.* 259:12724-12732). Mutations at this site promote attP, attB intramolecular recombination in mammalian cells (Lorbach *et al.* (2000) *J. Mol. Biol* 296:1175-1181).

To construct nucleic acid encoding the mutant, lambda integrase was PCR amplified from bacteriophage lambda DNA (cl857 *ind Sam* 7; New England Biolabs) using the following primers:

Lamint1 (SEQ ID No. 3)

-109-

TTCGAATTCATGGGAAGAAGGCGAAGTCATGAGCG)

Lamint2 (SEQ ID No. 4)

(TTCGAATTCTTATTTGATTTCAATTTTGTCCAC).

The resulting PCR product was digested with *EcoR* I and cloned into the  
5 *EcoR* I site of pUC19. Lambda integrase was mutated at amino acid  
position 174 using QuikChange Site-Directed Mutagenesis Kit  
(Stratagene) and the following oligos (generating a glutamic acid to  
arginine change at position 174):

LambdaINTE174R

10 (SEQ ID No. 6)

(CGCGCAGCAAAATCTAGAGTAAGGAGATCAAGACTTACGGCTGACG),

LamintR174rev (SEQ ID No. 7)

(CGTCAGCCGTAAGTCTTGATCTCCTTACTCTAGATTTTGCTGCGCG).

The resulting site directed mutant was confirmed by sequence analysis.

15 The wildtype and mutant lambda genes were cloned into the *EcoR* I site  
of pCX creating pCX-LamInt (SEQ ID NO: 127) and pCXLamIntR (Figure  
8; SEQ ID NO: 112).

The plasmid pCX (SEQ ID No. 70) was derived from plasmid  
pCXeGFP (SEQ ID No. 71). Excision of the *EcoRI* fragment containing the  
20 eGFP marker generated pCX. To generate plasmid pCXLamINTR (SEQ ID  
NO: 112) an *EcoRI* fragment containing the lambda integrase E174R (SEQ  
ID No. 37) mutation was cloned into the *EcoRI* site of pCX, and to  
generate plasmid pCX-LamINT, an *EcoRI* fragment containing the wild-  
type lambda integrase was cloned into the *EcoRI* site of pCX.

25 **B. Construction of integration vector containing attB and DsRed**

The plasmid pDsRedN1 (Clontech Laboratories, Palo Alto, CA; SEQ  
ID No. 29) was digested with *Hpa* I and ligated to the following annealed  
oligos:

attB1 (SEQ ID No. 8)

-110-

(TGAAGCCTGCTTTTTTATACTAACTTGAGCGAA)

attB2 (SEQ ID No. 9)

(TTCGCTCAAGTTAGTATAAAAAAGCAGGCTTCA)

The resulting vector (pDsRedN1-attB) was confirmed by PCR and  
5 sequence analysis.

**C. Transfection into LMtk- cells**

LM(tk-) cells containing the Prototype A ACes (L1-18; Chromos  
Molecular Systems Inc., Burnaby, BC Canada) were co-transfected with  
pDsRedN1 or pDsRedN1-attB and either pCXLamInt (SEQ ID NO: 127) or  
10 pCXLamIntR (SEQ ID NO: 112) using Lipofectamine Plus Reagent  
(LifeTechnologies, Gaithersburg, MD). The transfected cells were grown  
in DMEM (LifeTechnologies, Gaithersburg, MD) with 10% FBS (CanSera)  
and G418 (CalBiochem) at a concentration of 1 mg/ml.

**D. Enrichment by cell sorting**

15 The transfected cells were sorted using a FACs Vantage SE cell  
sorter (Becton Dickenson) to enrich for cells expressing DsRed. The cells  
were excited with a 488 nm Argon laser at 200 watts and cells  
fluorescing in the 585/42 detection channel were collected. The sorted  
cells were returned to growth medium for recovery and expansion. After  
20 three successive enrichments for cells expressing DsRed, single cell  
sorting into 96 well plates was performed using the same parameters.  
Duplicate plates of the single cell clones were made for PCR analysis.

**E. PCR analysis of single cell clones**

25 Pools of cells from each row and column of the 96 well plate were  
used for DNA isolation. DNA was prepared using a Wizard Genomic DNA  
purification kit (Promega Inc, Madison, WI). Nested PCR analysis on the  
DNA pools was performed to confirm the site-specific recombination  
event using the following primer sets:

-111-

attPdwn2 (SEQ ID No. 10)  
(TCTTCTCGGGCATAAGTCGGACACC)

CMVen (SEQ ID No. 11)  
(CTCACGGGGATTTCCAAGTCTCCAC)

5 followed by:

attPdwn (SEQ ID No. 12)  
(CAGAGGCAGGGAGTGGGACAAAATTG)

CMVen2 (SEQ ID No. 13)  
(CAACTCCGCCCCATTGACGCAAATG).

10 The resulting PCR reactions were analyzed by gel electrophoresis and the potential individual clones containing the site-specific recombination event were identified by combining the PCR results of all of the pooled rows and columns for each 96 well plate. The individual clones were then further analyzed by PCR using the following primers that flank the  
15 recombination junction. L1for and F1rev flank the attR junction whereas REDfor and L2rev flank the attL junction (see Figure 7):

L1for (SEQ ID No. 14)  
AGTATCGCCGAACGATTAGCTCTTCA

20 F1rev (SEQ ID No. 15)  
GCCGATTTTCGGCCTATTGGTTAAA

REDfor (SEQ ID No. 16)  
CCGCCGACATCCCCGACTACAAGAA

L2rev (SEQ ID No. 17)  
TTCCTTCGAAGGGGATCCGCCTACC.

25 **F. Sequence analysis of recombination junctions**

PCR products spanning the recombination junction were Topo-cloned into pcDNA3.1D/V5His (Invitrogen Inc., San Diego, CA) and then sequenced by cycle-sequencing. The clones were confirmed to have the correct *attR* and *attL* junctions by cycle sequencing.

30 **G. Fluorescent In Situ Hybridization (FISH)**

The cell lines containing the correct recombination junction sequence were further analyzed by fluorescent *in situ* hybridization (FISH)

-112-

by probing with the DsRed coding region labeled with biotin and visualizing with the Tyramide Signal Amplification system (TSA; NEN Life Science Products). The results indicate that the RFP sequence is present on the ACes.

5           **H. Southern analysis**

Genomic DNA was harvested from the cell lines containing an ACes with the correct recombinant event and digested with *EcoR* I. The digested DNAs were separated on a 0.7% agarose gel, transferred and fixed to a nylon membrane and probed with RFP coding sequences. The  
10 result showed that there is an integrated copy of RFP coding sequence in each clone.

**EXAMPLE 5**

**Delivery of a second gene encoding GFP onto the RFP platform ACes**

15           **A. Construction of integration vector containing attB and GFP (pD2eGFPiresPuroattB).**

The plasmid pRESpuro2 (Clontech, Palo Alto, CA; SEQ ID NO: 88) was digested with *EcoRI* and *NotI* then ligated to the D2eGFP *EcoRI-NotI* fragment from pD2eGFP-N1 (Clontech, Palo Alto, CA) to create pD2eGFPiresPuro2. Subsequently, oligos encoding the attB site were  
20 annealed and ligated into the *NruI* site of pD2eGFPiresPuro2 to create pD2eGFPiresPuroattB. The orientation of attB in the *NruI* site was determined by PCR.

**B. Transfection of LMtk- cells**

The LMtk- cells containing the RFP platform ACes produced in  
25 Example 4, which has multiple attP sites, were co-transfected with pCXLamIntR and pD2eGFPiresPuroattB using LipofectAMINE PLUS reagent. Five  $\mu$ g of each vector was placed into a tube containing 750  $\mu$ l of DMEM (Dulbecco's modified Eagles Medium). Twenty  $\mu$ l of the Plus reagent was added to the DNA and incubated at room temperature for 15

-113-

minutes. A mixture of 30  $\mu$ l of lipofectamine and 750  $\mu$ l DMEM was added to the DNA mixture and incubated an additional 15 minutes at room temperature. The DNA mixture was then added dropwise to approximately 3 million cells attached to a 10cm dish in 5 mls of DMEM.

- 5 The cells were incubated 4 hours (37°C, 5% CO<sub>2</sub>) with the DNA-lipid mixture, after which DMEM with 20% fetal bovine serum was added to the dishes to bring the culture medium to 10% fetal bovine serum. The dishes were incubated at 37°C with 5% CO<sub>2</sub>.

- Plasmid pD2eGFPiresPuroattB has a puromycin gene transcriptionally linked to the GFP gene *via* an IRES element. Two days after the transfection the cells were placed in medium containing puromycin at 4 $\mu$ g/ml to select for cells containing the pD2eGFPiresPuroattB plasmid integrated into the genome. Twenty-three clones were isolated after 17 days of selection with puromycin. These clones were expanded and then analyzed for the presence of the GFP gene on the ACes by 2-color (RFP/biotin & GFP/digoxigenin) TSA-FISH (NEN) according to the manufacturers protocol. Sixteen of the 23 clones produced a positive FISH signal on the ACes with a GFP probe.

#### EXAMPLE 6

#### 20 Delivery Of ACes Into human Mesenchymal Stem Cells (hMSC)

##### A. Transfection

- Transfection conditions for the most efficient delivery of the ACes into hMSCs (Cambrex BioWhittaker Product Code PT-2501, lot# F0658, East Rutherford, New Jersey) were assayed using LipofectAMINE PLUS and Superfect. One million prototype B ACes, which is a murine derived 60Mb ACes having primarily murine pericentric heterochromatin, and carrying a "payload" containing a hygromycin B selectable marker gene and a *lacZ* reporter gene (see , Telenius et al., 1999, Chrom. Res., 7:3-7; and Kereso et al., 1996, Chrom. Res., 4:226-239; each of which is



-114-

incorporated herein by reference in its entirety), were combined with 1-12  $\mu$ l of the transfection agent. In the case of LipofectAMINE PLUS, the PLUS reagent was combined with the ACes for 15 minutes followed by LipofectAMINE for a further 15 minutes. Superfect was complexed for 5 10 minutes at a ratio of 2  $\mu$ l Superfect per 1 million ACes. The ACes/transfection agent complex was then applied to 0.5 million recipient cells and the transfection was allowed to proceed according to the manufacturer's protocol. Percent transfected cells was determined on a FACS Vantage flow cytometer with argon laser tuned to 488 nm at 10 200mW and FITC fluorescence collected through a standard FITC 530/30 nm band pass filter. After 24 hours, IdUrd labeled ACes were delivered to human MSCs in the range of 30-50%, varying with transfection agent and dose. ACes delivery curves were generated from data collected in experiments that varied the dose of the transfection reagents. Dose 15 response curves of Superfect and LipofectAMINE PLUS, showing delivery of ACes into recipient hMSCs cells, were prepared, measured by transfer of IdUrd labeled ACes and detected by flow cytometry. Superfect shows maximum delivery in the range of 30-50% at doses greater than 2  $\mu$ l per million ACes. LipofectAMINE PLUS has a 42-48% delivery peak around 20 5-8  $\mu$ l per million ACes. These dose curves were then correlated with toxicity data to determine the transfection conditions that will allow for highest potential transfection efficiency. Toxicity was determined by a modified plating efficiency assay (de Jong et al., 2001, Chrom. Research, 9:475-485). The population's normalized plating efficiency (at maximum 25 % delivery doses) was in the range of 0.2 - 0.4 for Superfect and 0.5 - 0.6 with LipofectAMINE PLUS.

Due to the transfected population consisting of mixed cell types, flow cytometry allowed for the assessment of ACes delivery into each sub-population and the purification of the target population. Flow profiles

-115-

showing forward scatter (cell size) and side scatter (internal cell granularity) revealed three distinct hMSC populations that were gated into three regions: R3 (small cell region), R4 (medium cell region), R5 (large cell region). Transfection conditions were further optimized by re-analyzing delivery curves and assessing the differences in delivery to each sub-population. Dose response curves of Superfect and LipofectAMINE were prepared showing % delivery to each sub-population represented by the gating on basis of cell size and granularity properties of the mixed population. Three distinct hMSC populations were gated and % delivery dose curves generated. Using Superfect and LipofectAMINE PLUS the overall % delivery increased with cell size (80-90% delivery in large cells). LipofectAMINE PLUS at high doses (8-12  $\mu$ l per 1 million ACes) shows an increase in the overall proportion of chromosome transfer to the small population (10-20%). This suggests an advantage to using this transfection agent if the small-undifferentiated cell population is the desired target host cell.

#### **B. Expression from Genes on ACes IN hMSCs**

Following the delivery screening process conducted in section (A) above, the most promising results were subjected to further analyses to monitor expression and verify the presence of structurally intact ACes. The transfection conditions employed for these experiments were exactly the same as those that had been used during the screening process. Short-term expression was monitored by transfecting hMSCs with ACes containing a RFP gene (red fluorescent protein) set forth in Example 2C as "D11C4". The unselected population was harvested at 72-96 hours post transfection and % positive fluorescent cells measured by flow cytometry. RFP expression was in the range of 1-20%.

Long term-gene expression was assayed by selecting for hygromycin B resistant cells over a period of 7-10 days. Cytogenetic

-116-

analysis was done to detect presence of intact *ACes* by Fluorescent *In Situ* hybridization (FISH), where metaphase chromosomes were hybridized to a mouse major satellite-DNA probe (targeting murine pericentric heterochromatin) and a lambda probe (hybridizing to the *lacZ* gene). The human mesenchymal transfected culture could not undergo standard sub-cloning as diffuse colonies form with limited doublings available for expansion. Cytogenetic analysis was performed on the entire population, sampling over a period of 3-10 days post-transfection. The hygromycin resistant population was then blocked in mitosis with colchicine and analyzed for presence of intact *ACes* by FISH. Preliminary FISH results show approximately 2-8% of the hMSC-transfected population had an intact *ACes*. This compared to rat skeletal muscle myoblast clones, which were in the range of 60-95%. To increase the % of intact *ACes* in the hMSC-transfected population an enrichment step can be utilized as described in Example 2C.

### C. Differentiation of The hMSCs

In initial experiments where transfected hMSCs cells have been induced to differentiate into adipose or osteocytes, the results indicate that the transfected cells appear to be differentiating at a rate comparable to the untransfected controls and the cultures are lineage specific as tested by microscopic examination, FISH, Oil Red O staining (adipocyte assay), and calcium secretion (osteocyte assay).

Accordingly, these results indicate that the artificial chromosomes (*ACes*) provided herein can be successfully transferred into hMSC target cells. Targeting MSCs (such as hMSCs) permits gene transfer into cells in an undifferentiated state where the cells are easier to expand and purify. The genetically modified cells can then be differentiated *in vitro* or injected into a site *in vivo* where the microenvironment will induce transformation into specific cell lineages.

-117-

**EXAMPLE 7****Delivery of a Promoterless Marker Gene to a Platform ACes**

Platform ACes containing pSV40attPsensePURO (Figure 4) were constructed as set forth in Examples 3 and 4.

**5 A. Construction of Targeting Vectors.**

The base vector p18attBZeo (3166bp; SEQ ID NO: 114) was constructed by ligating the 1067bp *HindIII*-*SspI* fragment containing attBZeo, obtained from pLITattBZeo (SEQ ID NO:91), into pUC18 (SEQ ID NO: 122) digested with *HindIII* and *SspI*.

10 1. p18attBZEO-eGFP (6119bp; SEQ ID NO: 126) was constructed by inserting the 2977bp *SpeI*-*HindIII* fragment from pCXeGFP (SEQ ID NO:71; Okabe, *et al.* (1997) *FEBS Lett* 407:313-319) containing the eGFP gene into p18attBZeo (SEQ ID NO: 114) digested with *HindIII* and *XbaI*.

15 2. p18attBZEO-5'6XHS4eGFP (Figure 10; 7631bp; SEQ ID NO: 116) was constructed by ligating the 4465bp *HindIII* fragment from pCXeGFPattB(6XHS4)2 (SEQ ID NO: 123) which contains the eGFP gene, under the regulation of the chicken beta actin promoter, 6 copies of the HS4 core element located 5' of the chicken beta actin promoter and the polyadenylation signal into the *HindIII* site of p18attBZeo (SEQ ID NO: 20 114).

3. p18attBZEO-3'6XHS4eGFP (Figure 11; 7600bp; SEQ ID NO: 115) was created by removing the 5'6XHS4 element from p18attBZeo-(6XHS4)2eGFP (SEQ ID NO: 110). p18attBZeo-(6XHS4)2eGFP was digested with *EcoRV* and *SpeI*, treated with Klenow and religated to form 25 p18attBZeo3'6XHS4eGFP (SEQ ID NO: 115).

4. p18attBZEO-(6XHS4)2eGFP (Figure 12; 9080bp; SEQ ID NO: 110) was created in two steps. First, the *EcoRI*-*SpeI* fragment from pCXeGFPattB(6XHS4)2 (SEQ ID NO: 123) which contains 6 copies of the HS4 core element was ligated into p18attBZeo (SEQ ID NO: 114)

-118-

digested with *EcoRI* and *XbaI* to create p18attBZeo6XHS4 (4615bp; SEQ ID NO: 117). Next, p18attBZeo6XHS4 was digested with *HindIII* and ligated to the 4465bp *HindIII* fragment from pCXeGFPattB(6XHS4)2 which contains the eGFP gene, under the regulation of the chicken beta actin promoter, 6 copies of the HS4 core element located 5' of the chicken beta actin promoter and the polyadenylation signal.

Table 2

Targeting plasmid	No. zeocin resistant clones	No. clones with expected PCR product size	No. clones with correct sequence at recombination junction
p18attBZEO-eGFP	12	12	NT*
p18attBZEO-5'6XHS4eGFP	11	11	NT
p18attBZEO-3'6XHS4eGFP	11	11	NT
p18attBZEO-(6XHS4)2eGFP	9	9	4/4

\*NT = not tested

#### B. Transfection and Selection with Drug.

The mouse cell line containing the 2<sup>nd</sup> generation platform ACE, B19-38 (constructed as set forth in Example 3), was plated onto four 10cm dishes at approximately 5 million cells per dish. The cells were incubated overnight in DMEM with 10% fetal calf serum at 37°C and 5% CO<sub>2</sub>. The following day the cells were transfected with 5µg of each of the 4 vectors listed in Example 7.A. above and 5µg of pCXLamIntR (SEQ ID NO: 112), for a total of 10µg per 10cm dish. Lipofectamine Plus reagent was used to transfect the cells according to the manufacturers protocol. Two days post-transfection zeocin was added to the medium at 500ug/ml. The cells were maintained in selective medium until colonies formed. The colonies were then ring-cloned (see, e.g., McFarland, 2000, Methods Cell Sci, Mar;22(1):63-66).

#### C. Analysis of Clones (PCR, SEQUENCING).

-119-

Genomic DNA was isolated from each of the candidate clones with the Wizard kit (Promega) and following the manufacturers protocol. The following primer set was used to analyze the genomic DNA isolated from the zeocin resistant clones: 5PacSV40 -

- 5 CTGTTAATTAAGTGTGGAATGTGTG TCAGTTAGGGTG (SEQ ID NO:76);  
Antisense Zeo - TGAACAGGGTCACGTCGTCC (SEQ ID NO:77). PCR amplification with the above primers and genomic DNA from the site-specific integration of any of the 4 zeocin vectors would result in a 673bp PCR product.

- 10 As set forth in Table 2, of the 4 zeocin resistant candidate clones thusfar analyzed by PCR, all 4 exhibit the correct sequence for a site-specific integration event.

#### EXAMPLE 8

Integration of a PCR product by site-specific recombination.

- 15 In this example a gene is integrated onto the platform ACes by site-specific recombination without cloning said gene into a vector.

##### A. PCR PRIMER DESIGN.

- 20 PCR primers are designed to contain an attB site at the 5' end of one of the primers in the primer set. The remaining primers, which could be one or more than one primer, do not contain an attB site, but are complementary to sequences flanking the gene or genes of interest and any associated regulatory sequences. In first example, 2 primers (one containing an attB site) are used to amplify a selective gene such as puromycin.

- 25 In a second example as shown in Figure 13, the primer set includes primers 1 & 2 that amplify the GFP gene without amplification of an upstream promoter. Primer 1 contains the attB site at the 5' end of the oligo. Primers 3 & 4 are designed to amplify the IRES-blasticidin DNA sequences from the vector pIRESblasticidin. The 5' end of primer 3

-120-

contains sequences complementary to the 5' end of primer 2 such that annealing can occur between 5' ends of the two primers.

**B. PCR REACTION AND SUBSEQUENT LIGATION TO CREATE CIRCULAR MOLECULES FROM THE PCR PRODUCT**

5 In the first example set forth above in Section A, the two PCR primers are combined with a puromycin DNA template such as pPUR (Clontech), a heat stable DNA polymerase and appropriate conditions for DNA amplification. The resulting PCR product (attB-Puromycin) is then then purified and self-ligated to form a circular molecule.

10 In the second example set forth above in Section A, amplification of the GFP gene and IRES-blasticidin sequences is accomplished by combining primers 1 & 2 with DNA template pD2eGFP and primers 3 & 4 with template pIRESblasticidin under appropriate conditions to amplify the desired template. After initial amplification of the two products (attB-GFP  
15 & IRES-blasticidin) in separate reactions, a second round of amplification using both of the PCR products from the first round of amplification together with primers 1 and 4 amplifies the fusion product attB-GFP-IRES-blasticidin (Figure 13). This technique of using complementary sequences in primer design to create a fusion product is employed in *Saccharomyces*  
20 *cerevisiae* for allele replacement (Erdeniz *et al* (1997) *Gen Res* 7:1174-1183). The amplified product is then purified from the PCR reaction mixture by standard methods and ligated to form a circular molecule.

**C. INTRODUCTION OF PCR PRODUCT ONTO THE ACes USING A RECOMBINASE**

25 The circular PCR product is then be introduced to the platform ACes using the bacteriophage lambda integrase E174R. The introduction can be performed *in vivo* by transfecting the pCXLamIntR (SEQ ID NO: 112) vector encoding the lambda integrase mutant E174R together with the circularized PCR product into a cell line containing the platform ACE.

-121-

**D. SELECTION FOR MARKER GENE**

The marker gene (in this case either puromycin, blasticidin or GFP) is used to enrich the population for cells containing the proper integration event. A proper integration event in the second example (Figure 14)

- 5 juxtaposes a promoter residing on the platform ACes 5' to the attB-GFP-IRES-Blasticidin PCR product, allowing for transcription of both GFP and blasticidin. If enrichment is done by drug selection, blasticidin is added to the medium on the transfected cells 24-48 hours post-transfection. Selection is maintained until colonies are formed on the plates. If
- 10 enrichment is done by cell sorting, cells are sorted 2-4 days post-transfection to enrich for cells expressing the fluorescent marker (GFP in this case).

**E. ANALYSIS OF CLONES**

- Clonal isolates are analyzed by PCR, FISH and sequence analysis to
- 15 confirm proper integration events.

**EXAMPLE 9****Construction of a human platform ACes "ACE 0.1"****A. CONSTRUCTION OF THE TARGETING VECTOR pPACrDNA**

- Genome Systems (IncyteGenomics) was supplied with the primers
- 20 5'HETS (GGGCCGAAACGATCTCAACCTATT; SEQ ID NO:78), and 3'HETS (CGCAGCGGCCCTCCTACTC; SEQ ID NO:79), which were used to amplify a 538bp PCR product homologous to nt 9680-10218 of the human rDNA sequences (GenBank Accession No. U13369) and used as a probe to screen a human genomic P1AC (P1 Artificial Chromosome)
- 25 library constructed in the vector pCYPAC2 (Ioannou *et al.* (1994) *Nat. Genet.* 6(1): 84-89). Genome Systems clone #18720 was isolated in this screen and contains three repeats of human rDNA as assessed by restriction analysis. GS clone #18720, was digested with PmeI, a restriction enzyme unique to a single repeat of the human rDNA (45Kbp),



-122-

and then religated to form pPACrDNA (Figure 15). The insert in pPACrDNA was analyzed by restriction digests and sequence analysis of the 5' and 3' termini. The pPACrDNA, rDNA sequences are homologous to Genbank Accession #U13369, containing an insert of about 45 kB comprising a single repeat beginning from the end of one repeat at ~33980 (relative to the Genbank sequence) through the beginning of the next repeat up to approximately 35120 (the repeat offset from that listed in the GenBank file). Thus, the rDNA sequence is just over 1 copy of the repeat extending from 33980 (+/-10bp) to the end of the first repeat (43Kbp) and continuing into the second repeat to bp 35120 (+/-10bp).

#### **B. TRANSFECTION AND ACes FORMATION.**

Five hundred thousand MSU1.1 cells (Morgan et al., 1991, Exp. Cell Res., Nov;197(1):125-136; provided by Dr. Justin McCormick at Michigan State University) were plated per 6cm plate (3 plates total) and allowed to grow overnight. The cells were 70-80% confluent the following day. One plate was transfected with 15µg pPACrDNA (linearized with *Pme* I) and 2µg pSV40attPsensePuro (linearized with *Sca* I; see Example 3). The remaining plates were controls and were transfected with either 20µg pBS (Stratagene) or 20µg pSV40attBsensePuro (linearized with *Sca* I). All three plates were transfected using a CaPO<sub>4</sub> protocol.

#### **C. SELECTION OF PUROMYCIN RESISTANT COLONIES**

One day post-transfection the cells were "glycerol shocked" by the addition of PBS medium containing 10% glycerol for 30 seconds. Subsequently, the glycerol was removed and replaced with fresh DMEM. Four days post-transfection selective medium was added. Selective medium contains 1ug/ml puromycin. The transfection plates were maintained at 37°C with 5% CO<sub>2</sub> in selective medium for 2 weeks at which point colonies could be seen on the plate transfected with

-123-

pPACrDNA and pSV40attPsensePuro. The colonies were ring-cloned from the plate on day 17 post-selection and expanded in selective medium for analysis. Only two colonies (M2-2d & M2-2b) were able to proliferate in the selective medium after cloning. No colonies were seen on the control plates after 37 days in selective medium.

#### **D. ANALYSIS OF CLONES**

FISH analysis was performed on the candidate clones to detect ACes formation. Metaphase spreads from the candidate clones were probed in multiple probe combinations. In one experiment, the probes used were biotin-labeled human alphoid DNA (pPACrDNA) and digoxigenin-labeled mouse major DNA (pFK161) as a negative control. Candidate M2-2d was single cell subcloned by flow sorting and the candidate subclones were reanalyzed by FISH. Subclone 1B1 of M2-2d was determined to be a platform ACes and is also designated human Platform ACE 0.1.

#### **EXAMPLE 10**

##### **Site-specific integration of a marker gene onto a human platform ACE 0.1**

The promoterless delivery method was used to deliver a promoterless blasticidin marker gene onto the human platform ACes with excellent results. The human ACes platform with a promoterless blasticidin marker gene resulted in 21 of 38 blasticidin resistant clones displaying a PCR product of the expected size from the population co-transfected with pLIT38attBBSRpolyA10 and pCXLamIntR (Figure 8; SEQ ID NOs. 111 and 112). Whereas, the population transfected with pBlueScript resulted in 0 blasticidin resistant colonies.

#### **A. CONSTRUCTION OF pLIT38attB-BSRpolyA10 & pLIT38attB-BSRpolyA2.**

The vector pLITMUS 38 (New England Biolabs; U.S. Patent No. 5,691,140; SEQ ID NO: 119) was digested with *EcoRV* and ligated to

-124-

two annealed oligomers, which form an attB site (attB1 5'-TGAAGCCTGCTTTTTTATACTAACTTGAGCGAA-3' (SEQ ID NO:8); attB2 5'-TTCGCTCAAGTTAGTATAAAAAAGCAGGCTTCA-3'; SEQ ID NO:9). This ligation reaction resulted in the vector pLIT38attB (SEQ ID NO: 120).

- 5 The blasticidin resistance gene and SV40 polyA site was PCR amplified with primers: 5BSD (ACCATGAAAACATTTAACATTTCTCAACA; SEQ ID NO:80) and SV40polyA (TTTATTTGTGAAATTTGTGATGCTATTGC; SEQ ID NO:81) using pPAC4 (Frengen, E., *et al.* (2000) Genomics 68 (2), 118-126; GenBank Accession No. U75992) as template. The blasticidin-
- 10 SV40polyA PCR product was then ligated into pLIT38attB at the *Bam*HI site, which was Klenow treated following digestion with *Bam*HI. pLIT38attB-BSDpolyA10 (SEQ ID NO: 111) and pLIT38attB-BSDpolyA2 (SEQ ID NO: 121) are the two resulting orientations of the PCR product ligated into the vector.

15 **B. TRANSFECTION OF MSU1.1 CELLS CONTAINING HUMAN PLATFORM ACE 0.1.**

- MSU1.1 cells containing human platform ACE 0.1 (see Example 9) was expanded and plated to five 10cm dishes with  $1.3 \times 10^6$  cells per dish. The cells were incubated overnight in DMEM with 10% fetal bovine
- 20 serum, at 37°C and 5% CO<sub>2</sub>. The following day the cells were transfected with 5µg of each plasmid as set forth in Table 3, for a total of 10µg of DNA per plate of cells transfected (see Table 3) using ExGen 500 *in vitro* transfection reagent (MBI fermentas, cat. no. R0511). The transfection was performed according to the manufacturers protocol.
- 25 Cells were incubated at 37°C with 5% CO<sub>2</sub> in DMEM with 10% fetal bovine serum following the transfection.

-125-

Table 3

Plate #	Plasmid 1	Plasmid 2	No. Bsd <sup>R</sup> Colonies
1	pBS	None	0
2	pCXLamInt	pLIT38attB-BSRpolyA10	16
3	pCXLamIntR	pLIT38attB-BSRpolyA10	40
4	pCXLamInt	pLIT38attB-BSRpolyA2	28
5	pCXLamIntR	pLIT38attB-BSRpolyA2	36

#### 10 C. SELECTION OF BLASTICIDIN RESISTANT CLONES.

Three days following the transfection the cells were split from a 10 cm dish to two 15cm dishes. The cells were maintained in DMEM with 10% fetal bovine serum for 4 days in the 15 cm dishes. Seven days post-transfection blasticidin was introduced into the medium. Stably transfected cells were selected with 1 $\mu$ g/ml blasticidin. The number of colonies formed on each plate is listed in Table 3. These colonies were ring-cloned and expanded for PCR analysis. Upon expansion in blasticidin containing medium some clones failed to live and therefore do not have corresponding PCR data.

#### 20 D. PCR ANALYSIS

Thirty-eight of the 40 clones from plate 3 grew after ring-cloning. Genomic DNA was isolated from these clones with the Promega Wizard Genomic cDNA purification kit, digested with *Eco*RI and used as template in a PCR reaction with the following primers: 3BSP – TTAATTTCGGG TATATTTGAGTGGA (SEQ ID NO:82); 5PacSV40 – CTGTTAATTAAGTGTGGAA TGTGTGTCAGTTAGGGTG (SEQ ID NO:76). The PCR conditions were as follows. 100ng of genomic DNA was

-126-

amplified with 0.5ul Herculase polymerase (Stratagene) in a 50ul reaction that contained 12.5pmole of each primer, 2.5mM of each dNTP, and 1X Herculase buffer (Stratagene). The reactions were placed in a PerkinElmer thermocycler programmed as follows: Initial denaturation at 95°C for 10 minutes; 35 cycles of 94°C for 1 minute, 53°C for 1 minute, 72°C for 1 minute, and 72°C for 1 minute; Final extension for 10 minutes at 72°C; and 4°C hold. If pLIT38attB-BSRpolyA10 integrates onto the human platform ACE 0.1 correctly, PCR amplification with the above primers should yield an 804bp product. Twenty-one of the 38 clones from plate 3 produced a PCR product of the expected 804bp size.

#### EXAMPLE 11

##### **Delivery of a Vector comprising a Promoterless Marker Gene and a gene encoding a therapeutic product to a Platform ACes**

Platform ACes containing pSV40attPsensePURO (Figure 4) were constructed as set forth in Examples 3 and 4.

#### **A. CONSTRUCTION OF DELIVERY VECTORS**

##### **1. Erythropoietin cDNA vector, p18EPOcDNA.**

The erythropoietin cDNA was PCR amplified from a human cDNA library (E. Perkins *et al.*, 1999, *Proc. Natl. Acad. Sci. USA* 96(5): 2204-2209) using the following primers: EPO5XBA - TATCTAGAATGGGGGTGC ACGAATGTCCTGCC (SEQ ID NO: 83); EPO3BSI - TACGTACGTCATC TGTCCTGTCCTGCAGGC (SEQ ID NO: 84). The cDNA was amplified through two successive rounds of PCR using the following conditions: heat denaturation at 95°C for 3 minutes; 35 cycles of a 30 second denaturation (95°C), 30 seconds of annealing (60°C), and 1 minute extension (72°C); the last cycle is followed by a 7 minute extension at 72°C. BIO-X-ACT (BIOLINE) was used to amplify the erythropoietin cDNA from 2.5ng of the human cDNA library in the first round of amplification. Five  $\mu$ l of the first amplification product was used

-127-

as template for the second round of amplification. Two PCR products were produced from the second amplification with Taq polymerase (Eppendorf), each product was cloned into pCR2.1-Topo (Invitrogen) and sequenced. The larger PCR product contained the expected cDNA  
5 sequence for erythropoietin. The erythropoietin cDNA was moved from pTopoEPO into p18attBZeo(6XHS4)2eGFP (SEQ ID NO: 110). pTopoEPO was digested with BsiWI and XbaI to release a 588 bp EPO cDNA. BsrGI and BsiWI create compatible ends. The eGFP gene was removed from p18attBZeo(6XHS4)2eGFP by digestion with BsiWI and XbaI, the 8.3 Kbp  
10 vector backbone was gel purified and ligated to the 588 bp EPO cDNA to create p18EPOcDNA (SEQ ID NO: 124).

## 2. Genomic erythropoietin vector, p18genEPO.

The erythropoietin genomic clone was PCR amplified from a human genomic library (Clontech) using the following primers: GENEPO3BSI -  
15 CGTACGTCATCTGTCCCCT GTCCTGCA (SEQ ID NO: 85); GENEPO 5XBA -TCTAGAATGGGGGT GCACGGTGAGTACT (SEQ ID NO: 86). The reaction conditions for the amplification were as follows: heat denaturation for 3 minutes (95°C); 30 cycles of a 30 second denaturation (95°C), 30 seconds annealing (from 65°C decreasing 0.5°C per cycle to  
20 50°C), and 3 minutes extension (72°C); 15 cycles of a 30 second denaturation (95°C), 30 seconds annealing (50°C), and 3 minute extension (72°C); the last cycle is followed by a 7 minute extension at 72°C. The erythropoietin genomic PCR product (2147 bp) was gel purified and cloned into pCR2.1Topo to create pTopogenEPO. Sequence  
25 analysis revealed 2bp substitutions and insertions in the intronic sequences of the genomic clone of erythropoietin. A partial digest with XbaI and complete digest with BsiWI excised the erythropoietin genomic insert from pTopogenEPO. The resulting 2158 bp genomic erythropoietin fragment was ligated into the 8.3 Kbp fragment resulting from the

-128-

digestion of p18attBZeo(6XHS4)2eGFP (SEQ ID NO: 110) with XbaI and BsrGI to create p18genEPO (SEQ ID NO: 125).

#### **B. TRANSFECTION AND SELECTION WITH DRUG**

The erythropoietin genomic and cDNA genes were each moved  
5 onto the platform ACes B19-38 (constructed as set forth in Example 3) by  
co-transfecting with pCXLaIntR. Control transfections were also  
performed using pCXLaInt (SEQ ID NO: 127) together with either  
p18EPOcDNA (SEQ ID NO: 124) or p18genEPO (SEQ ID NO: 125).  
Lipofectamine Plus was used to transfect the DNA's into B19-38 cells  
10 according to the manufacturer's protocol. The cells were placed in  
selective medium (DMEM with 10% FBS and Zeocin @ 500ug/ml) 48  
hours post-transfection and maintained in selective medium for 13 days.  
Clones were isolated 15 days post-transfection.

#### **C. ANALYSIS OF CLONES (ELISA, PCR)**

##### **15 1. ELISA Assays**

Thirty clones were tested for erythropoietin production by an ELISA  
assay using a monoclonal anti-human erythropoietin antibody (R&D  
Systems, Catalogue # MAB287), a polyclonal anti-human erythropoietin  
antibody (R & D Systems, Catalogue # AB-286-NA) and alkaline  
20 phosphatase conjugated goat-anti-rabbit IgG (heavy and light chains)  
(Jackson ImmunoResearch Laboratories, Inc., Catalogue # 111-055-144).  
The negative control was a Zeocin resistant clone isolated from B19-38  
cells transfected with p18attBZeo(6XHS4) (SEQ ID NO: 117; no insert  
control vector) and pCXLaIntR (SEQ ID NO: 112). The preliminary  
25 ELISA assay was executed as follows: 1) Nunc-Immuno Plates (MaxiSorb  
96-well, Catalogue # 439454) were coated with 75ul of a 1/200 dilution  
(in Phosphate buffered Saline, pH 7.4 (PBS), Sigma Catalogue # P-3813)  
of monoclonal anti-human erythropoietin antibody overnight at 4°C. 2)  
The following day the plates were washed 3 times with 300ul PBS

-129-

containing 0.15% Tween 20 (Sigma, Catalogue # P-9416). 3) The plates were then blocked with 300ul of 1% Bovine Serum Albumin (BSA; Sigma Catalogue # A-7030) in PBS for 1 hour at 37°C. 4) Repeat the washes as in step 2. 5) The clonal supernatants (75ul per clone per well of 96-well plate) were then added to the plate and incubated for 1 hour at 37°C. The clonal supernatant analyzed in the ELISA assay had been maintained on the cells 7 days prior to analysis. 6) Repeat the washes of step 2. 7) Add 75ul of polyclonal anti-human erythropoietin antibody (1/250 dilution in dilution buffer (0.5% BSA, 0.01% Tween 20, 1X PBS, pH 7.4) and incubate 1 hour at 37°C. 8) Repeat washes of step 2. 9) Add 75ul of goat anti-rabbit conjugated alkaline phosphatase diluted 1/4000 in dilution buffer and incubate 1 hour at 37°C. 10) Repeat washes of step 2. 11) Add 75ul substrate, p-nitrophenyl phosphate (Sigma N2640), diluted to 1mg/ml in substrate buffer (0.1 Ethanolamine-HCl (Sigma, Catalogue # E-6133), 5mM MgCl<sub>2</sub> (Sigma, Catalogue # M-2393), pH 9.8). Incubate the plates in the dark for 1 hour at room temperature (22°C). 12) Read the absorption at 405nm (reference wavelength 495nm) on an Universal Microplate Reader (Bio-Tek Instruments, Inc., model # ELX800 UV). The erythropoietin standard curve was derived from readings of diluted human recombinant Erythropoietin (Roche, catalogue # 1-120-166; dilution range 125 - 7.8mUnits/ml). From this preliminary assay the 21 clones displaying the highest expression of erythropoietin were analyzed a second time in the same manner using medium supernatants that had been on the clones for 24 hours and a 1:3 dilution thereof.

## 25           2. PCR Analysis

Genomic DNA was isolated from the 21 clones with the best expression (as assessed by the initial ELISA assay above) as well as the B19-38 cell line and used for PCR analysis. Genomic DNA was isolated using the Wizard genomic DNA purification kit (Promega) according to the



-130-

manufacturers protocol. Amplification was performed on 100ng of genomic DNA as template with MasterTaq DNA Polymerase (Eppendorf) and the primer set 5PacSV40 – CTGTTAATTAAGTGTGGAATGTGTG TCAGTTAGGGTG (SEQ ID NO: 76) and Antisense Zeo -

- 5 TGAACAGGGTCACGTCGTCC (SEQ ID NO: 77). The amplification conditions were as follows: heat denaturation for 3 minutes (95°C); 30 cycles of a 30 second denaturation (95°C), 30 seconds annealing (from 65°C decreasing 0.5°C per cycle to 50°C), and 1 minutes extension (72°C); 15 cycles of a 30 second denaturation (95°C), 30 seconds
- 10 annealing (50°C), and 1 minute extension (72°C); the last cycle is followed by a 10 minute extension at 72°C. PCR products were size separated by gel electrophoresis. Of the 21 clones analyzed 19 produced a PCR product of 650 bp as expected for a site-specific integration event. All nineteen clones were the result of transformations with p19EPOcDNA
- 15 (5) or p18genEPO (14) and pCXLaIntR (i.e. mutant integrase). The remaining two clones, both of which were the result of transformation with p18genEPO (SEQ ID NO: 125) and pCXLaInt (i.e. wildtype integrase; SEQ ID NO: 127), produced a 400 bp PCR product.

#### Example 12

#### 20 Preparation of a Transformation Vector Useful for the Induction of Plant Artificial Chromosome Formation

- Plant artificial chromosomes (PACs) can be generated by introducing nucleic acid, such as DNA, which can include a targeting DNA, for example rDNA or lambda DNA, into a plant cell, allowing the cell
- 25 to grow, and then identifying from among the resulting cells those that include a chromosome with a structure that is distinct from that of any chromosome that existed in the cell prior to introduction of the nucleic acid. The structure of a PAC reflects amplification of chromosomal DNA, for example, segmented, repeat region-containing and heterochromatic

-131-

structures. It is also possible to select cells that contain structures that are precursors to PACs, for example, chromosomes containing more than one centromere and/or fragments thereof, and culture and/or manipulate them to ultimately generate a PAC within the cell.

5 In the method of generating PACs, the nucleic acid can be introduced into a variety of plant cells. The nucleic acid can include targeting DNA and/or a plant expressable DNA encoding one or multiple selectable markers (*e.g.*, DNA encoding bialophos (bar) resistance) or scorable markers (*e.g.*, DNA encoding GFP). Examples of targeting DNA  
10 include, but are not limited to, *N. tabacum* rDNA intergenic spacer sequence (IGS) and *Arabidopsis* rDNA such as the 18S, 5.8S, 26S rDNA and/or the intergenic spacer sequence. The DNA can be introduced using a variety of methods, including, but not limited to *Agrobacterium*-mediated methods, PEG-mediated DNA uptake and electroporation using,  
15 for example, standard procedures according to Hartmann *et al* [(1998) *Plant Molecular Biology* 36:741]. The cell into which such DNA is introduced can be grown under selective conditions and can initially be grown under non-selective conditions and then transferred to selective media. The cells or protoplasts can be placed on plates containing a  
20 selection agent to grow, for example, individual calli. Resistant calli can be scored for scorable marker expression. Metaphase spreads of resistance cultures can be prepared, and the metaphase chromosomes examined by FISH analysis using specific probes in order to detect amplification of regions of the chromosomes. Cells that have artificial  
25 chromosomes with functioning centromeres or artificial chromosomal intermediate structures, including, but not limited to, dicentric chromosomes, formerly dicentric chromosomes, minichromosomes, heterochromatin structures (*e.g.* sausage chromosomes), and stable self-replicating artificial chromosomal intermediates as described herein, are

-132-

identified and cultured. In particular, the cells containing self-replicating artificial chromosomes are identified.

The DNA introduced into a plant cell for the generation of PACs can be in any form, including in the form of a vector. An exemplary  
5 vector for use in methods of generating PACs can be prepared as follows.

For the production of artificial chromosomes, plant transformation vectors, as exemplified by pAgIIa and pAgIIb, containing a selectable marker, a targeting sequence, and a scorable marker were constructed using procedures well known in the art to combine the various fragments.  
10 The vectors can be prepared using vector pAg1 as a base vector and inserting the following DNA fragments into pAg1: DNA encoding  $\beta$ -glucoronidase under the control of the nopaline synthase (NOS) promoter fragment and flanked at the 3' end by the NOS terminator fragment, a fragment of mouse satellite DNA and an *N. tabacum* rDNA intergenic  
15 spacer sequence (IGS). In constructing plant transformation vectors, vector pAg2 can also be used as the base vector.

#### 1. Construction of pAG1

Vector pAg1 (SEQ. ID. NO: 89) is a derivative of the CAMBIA vector named pCambia 3300 (Center for the Application of Molecular  
20 Biology to International Agriculture, i.e., CAMBIA, Canberra, Australia; [www.cambia.org](http://www.cambia.org)), which is a modified version of vector pCambia 1300 to which has been added DNA from the bar gene conferring resistance to phosphinothricin. The nucleotide sequence of pCambia 3300 is provided in SEQ. ID. NO: 90. pCambia 3300 also contains a lacZ alpha sequence  
25 containing a polylinker region.

pAg1 was constructed by inserting two new functional DNA fragments into the polylinker of pCambia 3300: one sequence containing an attB site and a promoterless zeomycin resistance-encoding DNA flanked at the 3' end by a SV40 polyA signal sequence, and a second

-133-

sequence containing DNA from the hygromycin resistance gene (hygromycin phosphotransferase) conferring resistance to hygromycin for selection in plants. Although the zeomycin-SV40 polyA signal fusion is not expected to function in plant cells, it can be activated in mammalian  
5 cells by insertion of a functional promoter element into the attB site by site-specific recombination catalyzed by the Lambda att integrase. Thus, the inclusion of the attB-zeomycin sequences allows for evaluation of functionality of plant artificial chromosomes in mammalian cells by activation of the zeomycin resistance-encoding DNA, and provides an att  
10 site for further insertion of new DNA sequences into plant artificial chromosomes formed as a result of using pAg1 for plant transformation. The second functional DNA fragment allows for selection of plant cells with hygromycin. Thus, pAg1 contains DNA from the bar gene conferring resistance to phosphinothricin, DNA from the hygromycin resistance gene,  
15 both resistance-encoding DNAs under the control of a separate cauliflower mosaic virus (CaMV) 35S promoter, and the attB-promoterless zeomycin resistance-encoding DNA.

pAg1 is a binary vector containing *Agrobacterium* right and left T-DNA border sequences for use in *Agrobacterium*-mediated transformation  
20 of plant cells or protoplasts with the DNA located between the border sequences. pAg1 also contains the pBR322 Ori for replication in *E. coli*. pAg1 was constructed by ligating *HindIII*/*PstI*-digested p3300attBZeo with *HindIII*/*PstI*-digested pBSCaMV35SHyg as follows.

a. Generation of p3300attBZeo

25 Plasmid pCambia 3300 was digested with *PstI*/*EcoRI* and ligated with *PstI*/*StuI*-digested pLITattBZeo (the nucleotide sequence of pLITattBZeo is provided in SEQ. ID. NO: 91. (containing DNA encoding the zeocin resistance gene and an attB Integrase recognition sequence) to generate p3300attBZeo which contains an attB site, a promoterless

-134-

zeomycin resistance-encoding DNA flanked at the 3' end by a SV40 polyA signal, and a reconstructed *Pst*I site.

**b. Generation of pBSCaMV35SHyg**

A DNA fragment containing DNA encoding hygromycin phosphotransferase flanked by the CaMV 35S promoter and the CaMV 35S polyA signal sequence was obtained by PCR amplification of plasmid pCambia 1302 (GenBank Accession No. AF234298 and SEQ. ID. NO: 92). The primers used in the amplification reaction were as follows:

CaMV35SpolyA:

5'-CTGAATTAACGCCGAATTAATTCGGGGGATCTG-3' SEQ. ID. NO: 93  
CaMV35Spr:

5'-CTAGAGCAGCTTGCCAACATGGTGGAGCA-3' SEQ. ID. NO: 94

The 2100-bp PCR fragment was ligated with *Eco*RV-digested pBluescript II SK+ (Stratagene, La Jolla, CA, U.S.A.) to generate pBSCaMV35SHyg.

**c. Generation of pAg1**

To generate pAg1, pBSCaMV35SHyg was digested with *Hind*III/*Pst*I and ligated with *Hind*III/*Pst*I-digested p3300attBZeo. Thus, pAg1 contains the pCambia 3300 backbone with DNA conferring resistance to phophinothricin and hygromycin under the control of separate CaMV 35S promoters, an attB-promoterless zeomycin resistance-encoding DNA recombination cassette and unique sites for adding additional markers, *e.g.*, DNA encoding GFP. The attB site can be used as described hereing for the addition of new DNA sequences to plant artificial chromosomes, including PACs formed as a result of using the pAg1 vector, or derivatives thereof, in the production of PACs. The attB site provides a convenient site for recombinase-mediated insertion of DNAs containing a homologous att site.

**2. pAG2**

-135-

The vector pAg2 (SEQ. ID. NO: 95) is a derivative of vector pAg1 formed by adding DNA encoding a green fluorescent protein (GFP), under the control of a NOS promoter and flanked at the 3' end by a NOS polyA signal, to pAg1. pAg2 was constructed as follows. A DNA fragment  
5 containing the NOS promoter was obtained by digestion of pGEM-T-NOS, or pGEMEasyNOS (SEQ. ID. NO: 96), containing the NOS promoter in the cloning vector pGEM-T-Easy (Promega Biotech, Madison, WI, U.S.A.), with *Xba*I/*Nco*I and was ligated to an *Xba*I/*Nco*I fragment of pCambia 1302 containing DNA encoding GFP (without the CaMV 35S promoter) to  
10 generate p1302NOS (SEQ. ID. NO: 97) containing GFP-encoding DNA in operable association with the NOS promoter. Plasmid p1302NOS was digested with *Sma*I/*Bsi*WI to yield a fragment containing the NOS promoter and GFP-encoding DNA. The fragment was ligated with *Pme*I/*Bsi*WI-digested pAg1 to generate pAg2. Thus, pAg2 contains DNA  
15 from the bar gene conferring resistance to phosphinothricin, DNA conferring resistance to hygromycin, both resistance-encoding DNAs under the control of a cauliflower mosaic virus 35S promoter, DNA encoding kanamycin resistance, a GFP gene under the control of a NOS promoter and the attB-zeomycin resistance-encoding DNA. One of skill in  
20 the art will appreciate that other fragments can be used to generate the pAg1 and pAg2 derivatives and that other heterologous DNA can be incorporated into pAg1 and pAg2 derivatives using methods well known in the art.

### 3. pAgIIa and pAgIIb transformation vectors

25 Vectors pAgIIa and pAgIIb were constructed by inserting the following DNA fragments into pAg1: DNA encoding  $\beta$ -glucuronidase, the nopaline synthase terminator fragment, the nopaline synthase (NOS) promoter fragment, a fragment of mouse satellite DNA and an *N. tabacum*

-136-

rDNA intergenic spacer sequence (IGS). The construction of pAgIIa and pAgIIb was as follows.

An *N. tabacum* rDNA intergenic spacer (IGS) sequence (SEQ. ID. NO: 98; see also GenBank Accession No. YO8422; see also Borysyuk *et al.* (2000) *Nature Biotechnology* 18:1303-1306; Borysyuk *et al.* (1997) *Plant Mol. Biol.* 35:655-660; U.S. Patent Nos. 6,100,092 and 6,355,860) was obtained by PCR amplification of tobacco genomic DNA. The IGS can be used as a targeting sequence by virtue of its homology to tobacco rDNA genes; the sequence is also an amplification promoter sequence in plants. This fragment was amplified using standard PCR conditions (*e.g.*, as described by Promega Biotech, Madison, WI, U.S.A.) from tobacco genomic DNA using the primers shown below:

NTIGS-FI

5'- GTG CTA GCC AAT GTT TAA CAA GAT G- 3' (SEQ ID No. 99) and  
NTIGS-RI

5'-ATG TCT TAA AAA AAA AAA CCC AAG TGA C- 3' (SEQ ID No. 100)  
Following amplification, the fragment was cloned into pGEM-T Easy to give pIGS-I. A fragment of mouse satellite DNA (Msat1 fragment; GenBank Accession No. V00846; and SEQ ID No. 101) was amplified via PCR from pSAT-1 using the following primers:

MSAT-F1

5'- AAT ACC GCG GAA GCT TGA CCT GGA ATA TCG C -3'(SEQ ID No. 102) and

MSAT-RI

5'-ATA ACC GCG GAG TCC TTC AGT GTG CA T- 3' (SEQ ID No. 103)

This amplification added a *SacII* and a *HindIII* site at the 5' end and a *SacII* site at the 3' end of the PCR fragment. This fragment was then cloned into the *SacII* site in pIGS-1 to give pMIGS-1, providing a eukaryotic

-137-

centromere-specific DNA and a convenient DNA sequence for detection via FISH.

A functional marker gene containing a NOS-promoter:GUS:NOS terminator fusion was then constructed containing the NOS promoter  
5 (GenBank Accession No. U09365; SEQ ID No. 104), *E. coli*  $\beta$ -glucuronidase coding sequence (from the GUS gene; GenBank Accession No. S69414; and SEQ ID No. 105), and the nopaline synthase terminator sequence (GenBank Accession No. U09365; SEQ ID No. 107). The NOS promoter in pGEM-T-NOS was added to a promoterless GUS  
10 gene in pBlueScript (Stratagene, La Jolla, CA, U.S.A.) using *NotI*/*SpeI* to form pNGN-1, which has the NOS promoter in the opposite orientation relative to the GUS gene.

pMIGS-1 was digested with *NotI*/*SpeI* to yield a fragment containing the mouse major satellite DNA and the tobacco IGS which was  
15 then added to *NotI*-digested pNGN-1 to yield pNGN-2. The NOS promoter was then re-oriented to provide a functional GUS gene, yielding pNGN-3, by digestion and religation with *SpeI*. Plasmid pNGN-3 was then digested with *HindIII*, and the *HindIII* fragment containing the  $\beta$ -glucuronidase coding sequence and the rDNA intergenic spacer, along with the Msat  
20 sequence, was added to pAG-1 to form pAgIIa (SEQ ID NO: 108), using the unique *HindIII* site in pAg1 located near the right T-DNA border of pAg1, within the T-DNA region.

Another plasmid vector, referred to as pAgIIb, was also recovered, which contained the inserted *HindIII* fragment (SEQ ID NO: 108) in the  
25 opposite orientation relative to that observed in pAgIIa. Thus, pAgIIa and pAgIIb differ only in the orientation of the *HindIII* fragment containing the mouse major satellite sequence, the GUS DNA sequence and the IGS sequence. The nucleotide sequences of pAgIIa is provided in SEQ. ID. NOS: 109.



-138-

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

-139-

**WHAT IS CLAIMED IS:**

1. A eukaryotic chromosome comprising one or a plurality of *att* site(s), wherein:
  - an *att* site is heterologous to the chromosome; and
  - 5 an *att* site permits site-directed integration in the presence of lambda integrase.
2. The eukaryotic chromosome of claim 1, wherein the *att* sites are selected from the group consisting of *attP* and *attB* or *attL* and *attR*, or variants thereof.
- 10 3. The eukaryotic chromosome of claim 1 that is an artificial chromosome.
4. The eukaryotic chromosome of claim 1 that is an artificial chromosome expression system (*ACes*).
5. The eukaryotic chromosome of claim 4 that is predominantly  
15 heterochromatin.
6. The chromosome of claim 1 that is an artificial chromosome that contains no more than about 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.
7. The chromosome of claim 1 that is a plant chromosome.
- 20 8. The chromosome of claim 1 that is an animal chromosome.
9. The chromosome of claim 7 that is a plant artificial chromosome.
10. The chromosome of claim 8 that is an animal artificial chromosome.
- 25 11. The chromosome of claim 8 that is a mammalian chromosome.
12. The chromosome of claim 11 that is a mammalian artificial chromosome.

-140-

13. The chromosome of claim 6 that is an artificial chromosome expression system (ACes).

14. A platform artificial chromosome expression system (ACes) comprising one or a plurality of sites that participate in recombinase  
5 catalyzed recombination.

15. The ACes of claim 14 that contains one site.

16. The ACes of claim 14 that is predominantly heterochromatin.

17. The ACes of claim 14 that contains no more than about 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.

10 18. The ACes of claim 14 that is a plant ACes.

19. The ACes of claim 14 that is an animal ACes.

20. The ACes of claim 14 that is selected from a fish, insect, reptile, amphibian, arachnid or a mammalian ACes.

21. The ACes of claim 14 that is a fish ACes.

15 22. The artificial chromosome expression system (ACes) of claim 14, wherein the recombinase and site(s) are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase system of phage Mu, the Cin recombinase system, the Pin recombinase system of *E. coli* and the R/RS  
20 system of the pSR1 plasmid, or any combination thereof.

23. A method of introducing heterologous nucleic acid into a chromosome, comprising:

contacting a chromosome of any of claims 1 or 14 with a nucleic acid molecule comprising both the heterologous nucleic acid and a  
25 recombination site, in the presence of a recombinase that promotes recombination between the sites in the chromosome and in the nucleic acid molecule.

-141-

24. The method of claim 23, wherein the recombinase is selected from the group consisting of Cre, Gin, Cin, Pin, FLP, a phage integrase and R from the pSR1 plasmid.

25. The method of claim 23, wherein the nucleic acid molecule  
5 encodes a therapeutic protein, antisense nucleic acid, or comprises an artificial chromosome.

26. The method of claim 25, wherein the nucleic acid molecule comprises a yeast artificial chromosomes (YAC), a bacterial artificial chromosome (BAC) or an insect artificial chromosome (IAC).

10 27. A combination, comprising, the chromosome of claim 1 and a first vector comprising the cognate recombination site, wherein the cognate recombination site is a site that recombines with the site engineered into the chromosome.

28. The combination of claim 27, further comprising nucleic acid  
15 encoding a recombinase, wherein the nucleic acid is on a second vector or on the first vector, or on the ACes under an inducible promoter.

29. The combination of claim 28, wherein the recombinase and sites are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase  
20 system of phage Mu, the Pin recombinase system of *E. coli* and the R/RS system of the pSR1 plasmid, or any combination thereof.

30. The combination of claim 28, wherein a vector is the plasmid pCXLamIntR.

31. The combination of claim 27, wherein a vector is the plasmid  
25 pDsRedN1-attB.

32. A kit, comprising the combination of claim 27 and optionally instructions for introducing heterologous nucleic acid into the chromosome.

-142-

33. A method for introducing heterologous nucleic acid into a platform artificial chromosome, comprising:

(a) mixing an artificial chromosome comprising at least a first recombination site and a vector comprising at least a second  
5 recombination site and the heterologous nucleic acid;

(b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the artificial chromosome.

10 34. The method of claim 33, wherein the artificial chromosome is an ACes.

35. The method of claim 33, wherein said mixing step (a) is conducted in cells ex vivo.

15 36. The method of claim 33, wherein said mixing step (a) is conducted extracellularly in an in vitro reaction mixture.

37. The method of claim 33, wherein the at least one recombination protein is encoded by a bacteriophage selected from the group consisting of bacteriophage lambda, phi 80, P22, P2, 186, P4 and P1.

20 38. The method of claim 37, wherein the at least one recombination protein is encoded by bacteriophage lambda, or mutants thereof.

39. The method of claim 33, wherein at least one recombination protein is selected from the group consisting of Int, IHF, Xis and Cre,  $\gamma\delta$ ,  
25 Tn3 resolvase, Hin, Gin, Cin and Flp.

40. The method of claim 32, wherein the recombination sites are selected from the group consisting of att and lox P sites.

-143-

41. The method of claim 33, wherein the first and/or second recombination site contains at least one mutation that removes one or more stop codons.

42. The method of claim 33, wherein the first and/or second  
5 recombination site contains at least one mutation that avoids hairpin formation.

43. The method of claim 33, wherein the first and/or second recombination site comprises at least a first nucleic acid sequence selected from the group consisting of SEQ ID NOs:41-56:

- 10 a) RKYCWGCTTTYKTRTACNAASTSGB (m-att) (SEQ ID NO:41);  
b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:42);  
c) GTTCAGCTTCTKTRTACNAACTSGB (m-attR) (SEQ ID NO:43);  
d) AGCCWGCTTCTKTRTACNAAAGTSGB (m-attL) (SEQ ID NO:44);  
e) GTTCAGCTTTYKTRTACNAAAGTSGB (m-attP1) (SEQ ID NO:45);  
15 f) AGCCTGCTTTTTTGTACAAACTTGT (attB1) (SEQ ID NO:46);  
g) AGCCTGCTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:47);  
h) ACCCAGCTTCTTGTACAAACTTGT (attB3) (SEQ ID NO:48);  
i) GTTCAGCTTTTTTGTACAAACTTGT (attR1) (SEQ ID NO:49);  
j) GTTCAGCTTCTTGTACAAACTTGT (attR2) (SEQ ID NO:50);  
20 k) GTTCAGCTTCTTGTACAAAGTTGG (attR3) (SEQ ID NO:51);  
l) AGCCTGCTTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:52);  
m) AGCCTGCTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:53);  
n) ACCCAGCTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:54);  
o) GTTCAGCTTTTTTGTACAAAGTTGG (attP1) (SEQ ID NO:55);  
25 p) GTTCAGCTTCTTGTACAAAGTTGG (attP2, P3) (SEQ ID NO:  
56);

and a corresponding or complementary DNA or RNA sequence,  
wherein R=A or G, K=G or T/U, Y=C or T/U, W=A or T/U, N=A or C  
or G or T/U, S=C or G, and B=C or G or T/U; and

-144-

the core region does not contain a stop codon in one or more reading frames.

44. The method of claim 33, wherein the first and/or second recombination site comprises at least a first nucleic acid sequence  
5 selected from the group consisting of a mutated att recombination site containing at least one mutation that enhances recombinational specificity, a complementary DNA sequence thereto, and an RNA sequence corresponding thereto.

45. The method of claim 33, wherein the vector comprising the  
10 second site further encodes at least one selectable marker.

46. The method of claim 45, wherein the marker is a promoterless marker, which, upon recombination is under the control of a promoter and is thereby expressed.

47. The method of claim 46, wherein the first recombination site  
15 is attP and is in the sense orientation prior to recombination.

48. The method of claim 46, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic, fluorescent, or capable of being bound by an antibody and FACs sorted.

20 49. The method of claim 48, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase (hisD).

50. A cell comprising, the chromosome of claim 1.

25 51. The cell of claim 50, wherein the cell is a nuclear donor cell.

52. The cell of claim 50, wherein the cell is a stem cell.

53. The stem cell of claim 52, wherein said stem cell is human and is selected from the group consisting of a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell and an embryonic stem cell.

-145-

54. The cell of claim 50, wherein the cell is mammalian.
55. The cell of claim 54, wherein the mammal is selected from the group consisting of humans, primates, cattle, pigs, rabbits, goats, sheep, mice, rats, guinea pigs, hamsters, cats, dogs, and horses.
- 5 56. The cell of claim 50, wherein the cell is a plant cell.
57. A cell comprising the platform *ACes* of claim 14.
58. The cell of claim 57, wherein the cell is a nuclear donor cell.
59. The cell of claim 57, wherein the cell is a stem cell.
60. The stem cell of claim 59, wherein said stem cell is human
- 10 and is selected from the group consisting of a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell and an embryonic stem cell.
61. A human mesenchymal cell comprising an artificial chromosome.
62. The human mesenchymal cell of claim 61, wherein said
- 15 artificial chromosome is an *ACes*.
63. The human mesenchymal cell of claim 62, wherein the *ACes* is a platform-*ACes*.
64. A method for introducing heterologous nucleic acid into the mesenchymal cell of claim 63, comprising:
- 20 (a) introducing into the cell of claim 63, wherein the platform-*ACes* has a first recombination site, a vector comprising at least a second recombination site and the heterologous nucleic acid;
- (b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between
- 25 the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the platform-*ACes* within the mesenchymal cell.
65. A lambda-intR mutein comprising a glutamic acid to arginine change at position 174 of wild-type lambda-intR.



-146-

66. The lambda-intR mutein of claim 65, wherein the lambda-intR mutein comprises SEQ ID NO:37.

67. The method of claim 46 wherein the promoterless marker is transcriptionally downstream of the heterologous nucleic acid, wherein  
5 the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.

68. The method of claim 67, wherein the selectable marker and the heterologous nucleic acid are transcriptionally linked by the presence  
10 of a IRES between them.

69. The method of claim 68, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.

70. The method of claim 69, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase.  
15

71. The method of claim 67 further comprising expressing the  
20 heterologous protein and isolating the heterologous protein.

72. A method for producing a transgenic animal, comprising introducing a platform-ACes into an embryonic cell.

73. The method of claim 72, wherein the embryonic cell is a stem cell.

74. The method of claim 72, wherein the embryonic cell is in an embryo.  
25

75. The method of claim 72, wherein the platform-ACes comprises heterologous nucleic acid that encodes a therapeutic product.

-147-

76. The method of claim 72, wherein the transgenic animal is a fish, insect, reptile, amphibians, arachnid or mammal.

77. The method of claim 72, wherein the *ACes* is introduced by cell fusion, lipid-mediated transfection by a carrier system, microinjection,  
5 microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer.

78. A transgenic animal produced by the method of claim 72.

79. A cell line useful for making a library of *ACes*, comprising a multiplicity of heterologous recombination sites randomly integrated  
10 throughout the endogenous chromosomes.

80. A method of making a library of *ACes* comprising random portions of a genome, comprising introducing one or more *ACes* into the cell line of claim 79, under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of  
15 the heterologous recombination sites within the cell's chromosomal DNA; and isolating said multiplicity of *ACes*, thereby producing a library of *ACes* whereby multiple *ACes* have different portions of the genome within.

81. A library of cells useful for genomic screening, said library  
20 comprising a multiplicity of cells, wherein each cell comprises an *ACes* having a mutually exclusive portion of a chromosomal nucleic acid therein.

82. The library of cells of claim 81, wherein the cells of the library are from a different species than the chromosomal nucleic acid  
25 within the *ACes*.

83. A method of making one or more cell lines, comprising

a) integrating into endogenous chromosomal DNA of a selected cell species, a multiplicity of heterologous recombination sites,

-148-

b) introducing a multiplicity of *ACes* under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites integrated within the cell's endogenous chromosomal DNA;

5        c) isolating said multiplicity of *ACes*, thereby producing a library of *ACes* whereby a multiplicity of *ACes* have mutually exclusive portions of the endogenous chromosomal DNA therein;

d) introducing the isolated multiplicity of *ACes* of step c) into a multiplicity of cells, thereby creating a library of cells;

10       e) selecting different cells having mutually exclusive *ACes* therein and clonally expanding or differentiating said different cells into clonal cell cultures, thereby creating one or more cell lines.

84. The method of claim 23, wherein the nucleic acid molecule with a recombination site is a PCR product.

15       85. Method of claim 23 wherein the recombinase is a protein and the recombination event occurs in vitro.

86. The method of claim 33, wherein the vector is a PCR product comprising a second recombination site.

20       87. The lambda-intR mutein of claim 65, wherein the mutein further comprises an amino acid signal for nuclear localization.

88. The lambda-intR mutein of claim 65, wherein the mutein further comprises an epitope tag for protein purification.

89. A modified iron-induced promoter comprising SEQ ID NO:128.

25       90. A plasmid or expression cassette comprising the promoter of claim 89.

91. A vector, comprising:  
a recognition site for recombination; and

-149-

a sequence of nucleotides that targets the vector to an amplifiable region of a chromosome.

92. The vector of claim 91, wherein the amplifiable region comprises heterochromatic nucleic acid.

5 93. The vector of claim 91, wherein the amplifiable region comprises rDNA.

94. The vector of claim 93, wherein the rDNA comprises an intergenic spacer.

95. The vector of claim 91, further comprising nucleic acid  
10 encoding a selectable marker that is not operably associated with any promoter.

96. The vector of claim 91, wherein the chromosome is a mammalian chromosome.

97. The vector of claim 91, wherein the chromosome is a plant  
15 chromosome.

98. A cell of claim 57 that is a plant cell, wherein the ACes platform is a MAC.

99. The plant cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

20 100. The plant cell of claim 99, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

101. A cell of claim 57 that is an animal cell, wherein the ACes platform is a plant artificial chromosome (PAC).

25 102. The cell of claim 101 that is a mammalian cell.

103. The cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

104. The cell of claim 102, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

-150-

105. The cell of claim 104, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

106. A method, comprising:

- 5       introducing a vector of claim 91 into a cell;  
      growing the cells; and  
      selecting a cell comprising an artificial chromosome that comprises one or more repeat regions.

107. The method of claim 106, wherein sufficient portion of the  
10       vector integrates into a chromosome in the cell to result in amplification of chromosomal DNA.

108. The method of claim 106, wherein the artificial chromosome is an *ACes*.

109. A method for screening, comprising:

- 15       contacting a cell comprising a reporter *ACes* with test compounds or known compounds, wherein:

      the reporter *ACes* comprises one or a plurality of reporter constructs;

- a reporter construct comprises a reporter gene in operative linkage  
20       with a regulatory region responsive to test or known compounds; and  
      detecting any increase or decrease in signal output from the reporter, wherein a change in the signal is indicative of activity of the test or known compound on the regulatory region.

110. The method of claim 109, wherein the reporter is operatively  
25       linked to a promoter that controls expression of a gene in a signal transduction pathway, whereby activation or reduction in the signal indicates that the pathway is activated or down-regulated by the test compound.

-151-

111. The method of claim 109, wherein the reporter in the construct encodes drug resistance or encodes a fluorescent protein.

112. The method of claim 111, wherein the fluorescent protein is selected from the group consisting of red, green and blue fluorescent  
5 proteins.

113. The method of claim 109, wherein the *ACes* comprises a plurality of reporter-linked constructs, each with a different reporter, whereby the pathway(s) affected by the test compounds can be elucidated.

10 114. The method of claim 109, wherein a reporter is operatively linked to a promoter that is transcriptionally regulated in response to DNA damage, and the test compounds are genotoxicants.

115. The method of claim 114, wherein the DNA damage is induced by apoptosis, necrosis or cell-cycle perturbations.

15 116. The method of claim 114, wherein unknown compounds are screened to assess whether they are genotoxicants.

117. The method of claim 114, wherein the promoter is a cytochrome P450-profiled promoter.

20 118. The method of claim 114, wherein the cell is in a transgenic animal and toxicity is assessed in the animal.

119. The method of claim 109, wherein:

the cell is a patient cell sample; the patient has a disease;

the regulatory region is one targeted by a drug or drug regimen;

and

25 the method assesses the effectiveness of a treatment for the disease for the particular patient.

120. The method of claim 119, wherein the cell is a tumor cell.

121. The method of claim 109, wherein the cell is a stem cell or a progenitor cell, whereby expression of the reporter is operatively linked to

-152-

a regulatory region expressed in the cells to thereby identify stem cells or progenitor cell.

122. The method of claim 109, wherein the cell is in an animal;  
and the method comprises whole-body imaging to monitor expression of  
5 the reporter in the animal.

123. A reporter ACes comprises one or a plurality of reporter constructs, wherein the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

Fig. 1 Generation of ACes for Platform Chromosome Engineering

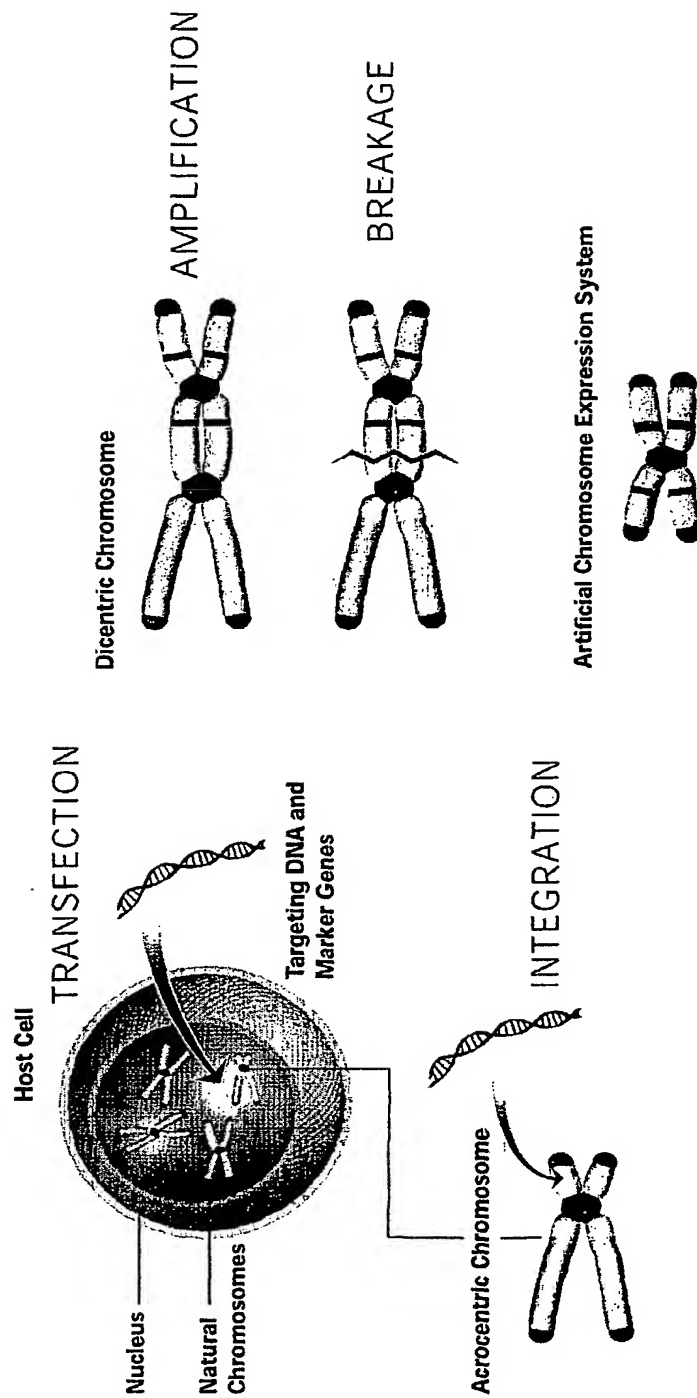
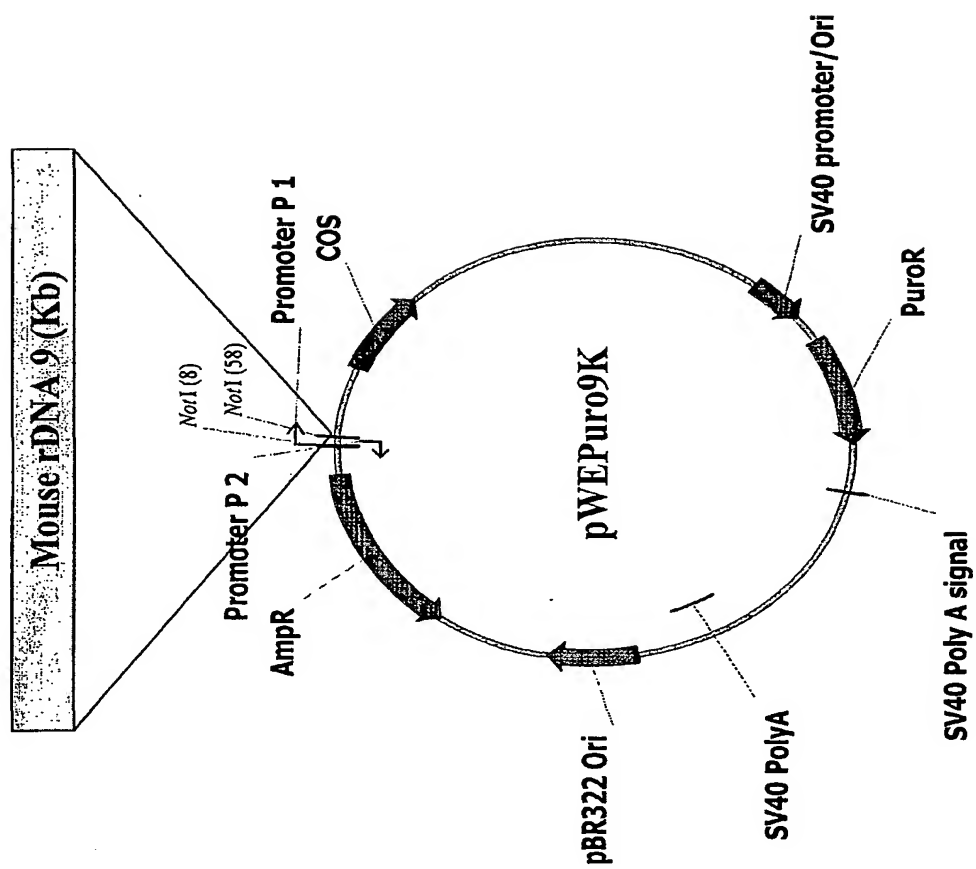




Fig. 2 Targeting plasmid pWEPuro9K



3/15

Fig. 3 Platform Chromosome with single recombination site

Platform chromosome containing mouse rDNA, mouse major/minor repeat DNA and puromycin resistant cassette



Targeting onto platform chromosome via homologous recombination



or



Engineered Selectable Marker System

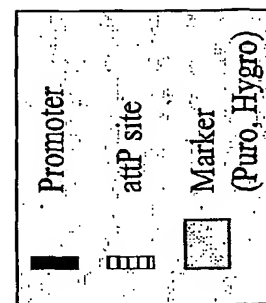
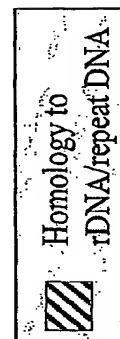
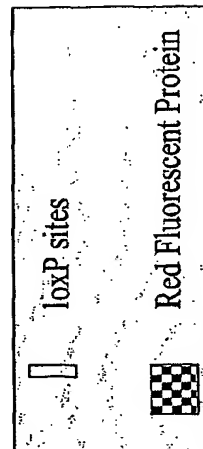
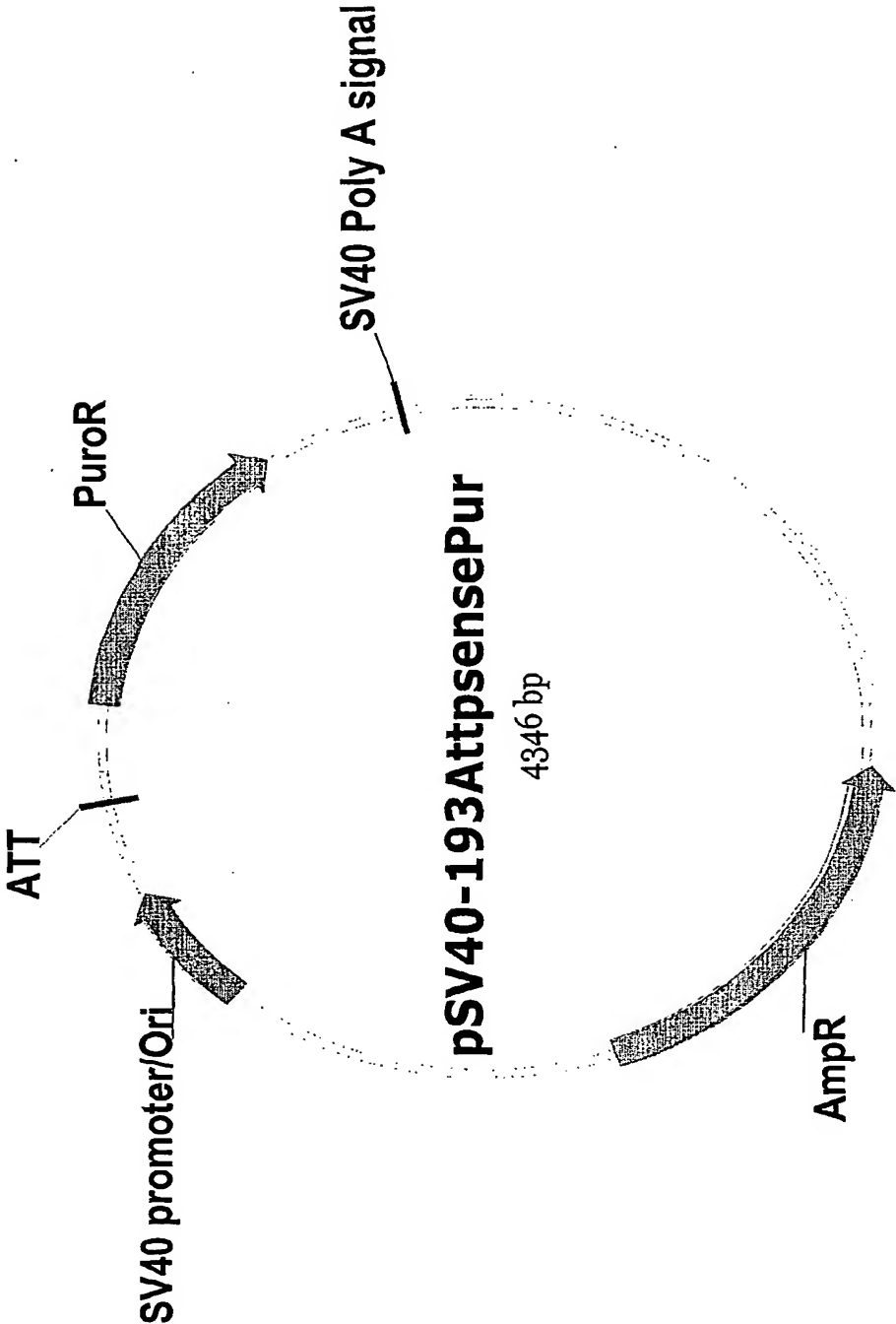


Fig. 4



5/15

Fig. 5 Generation of chromosome based platform with multiple integration sites

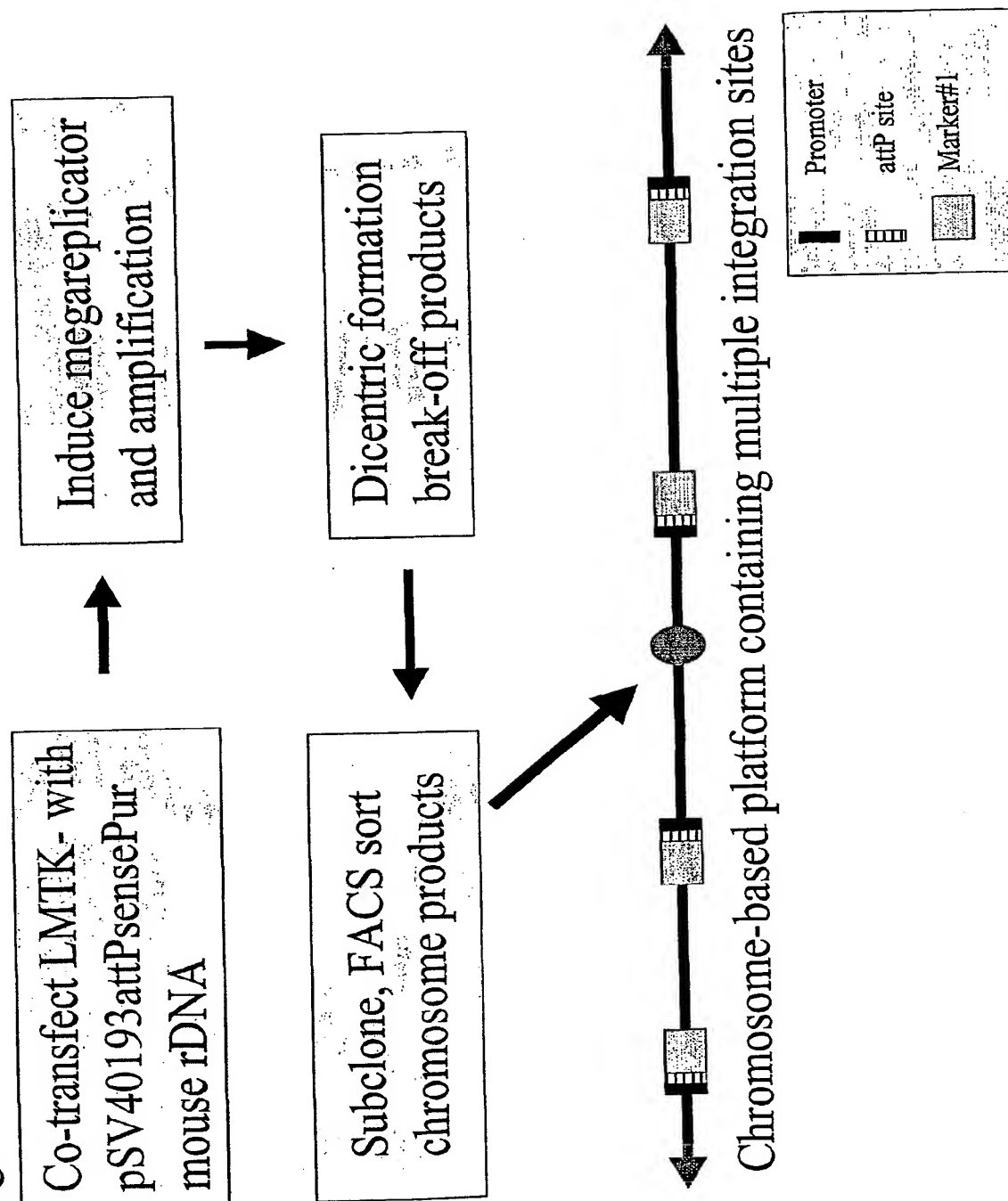


Fig. 6  $\lambda$  integrase recombination

Core Region

attP CAGCTTTTTTATACTAAGTTG

attB CTGCTTTTTTATACTAACTTG

attL CTGCTTTTTTATACTAAGTTG

attR CAGCTTTTTTATACTAACTTG

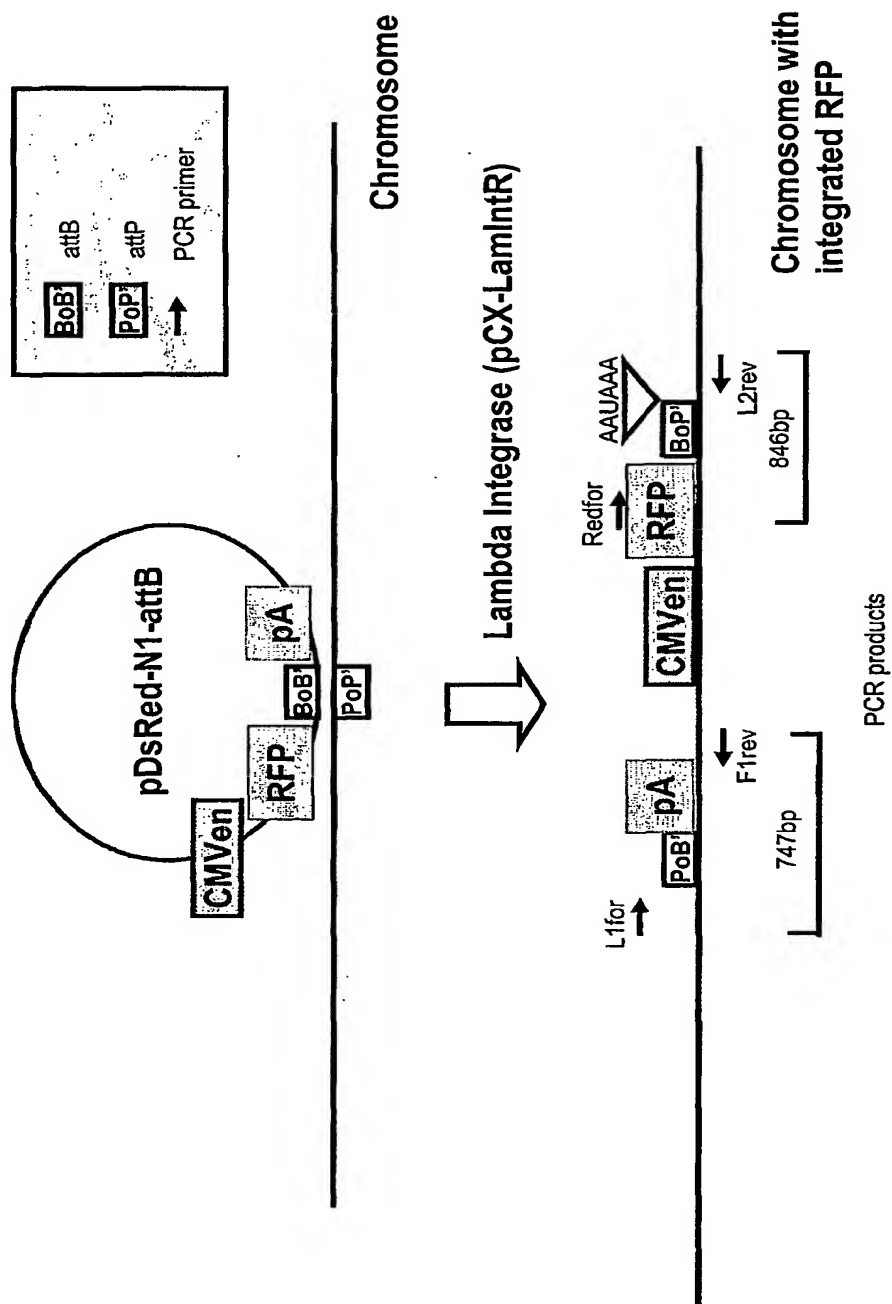
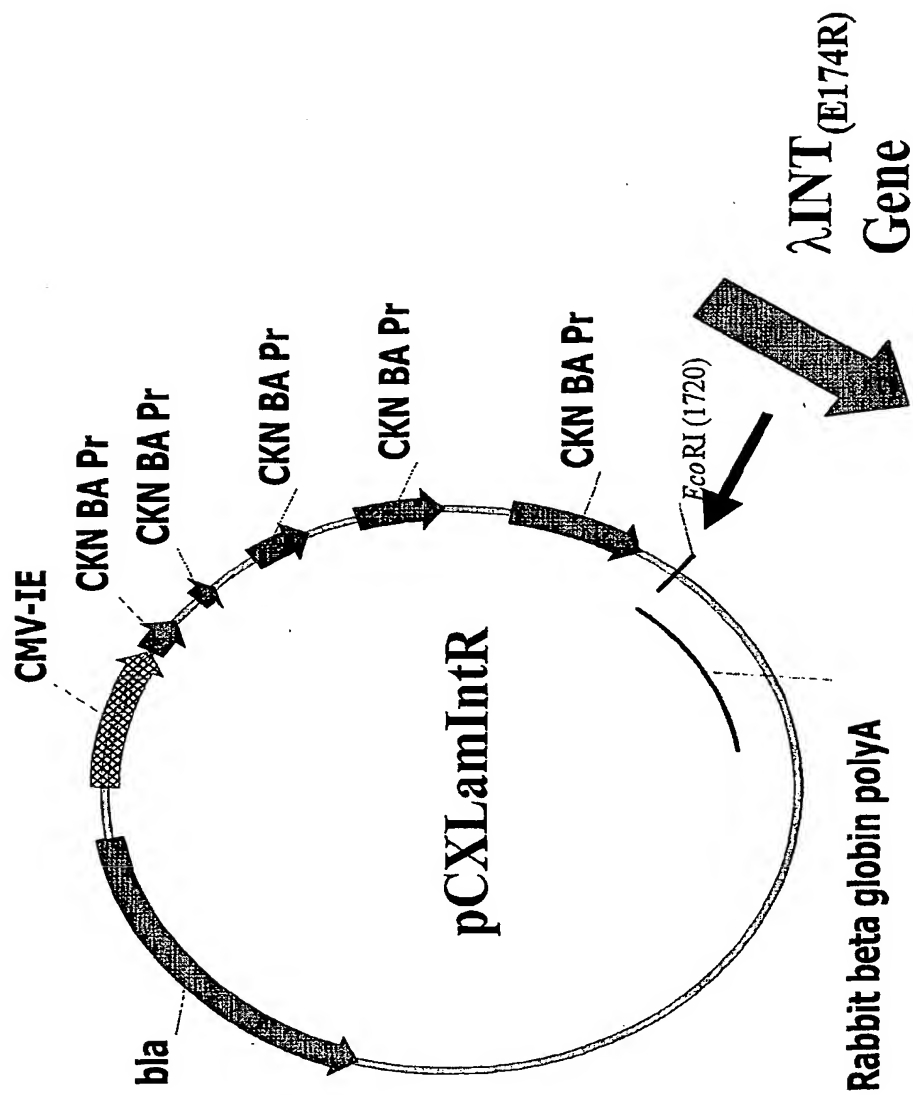
Fig. 7  $\lambda$ INT recombination on artificial chromosome

Fig. 8 pCXLamIntR integrase expression vector



9/15

Fig. 9 Platform Technology

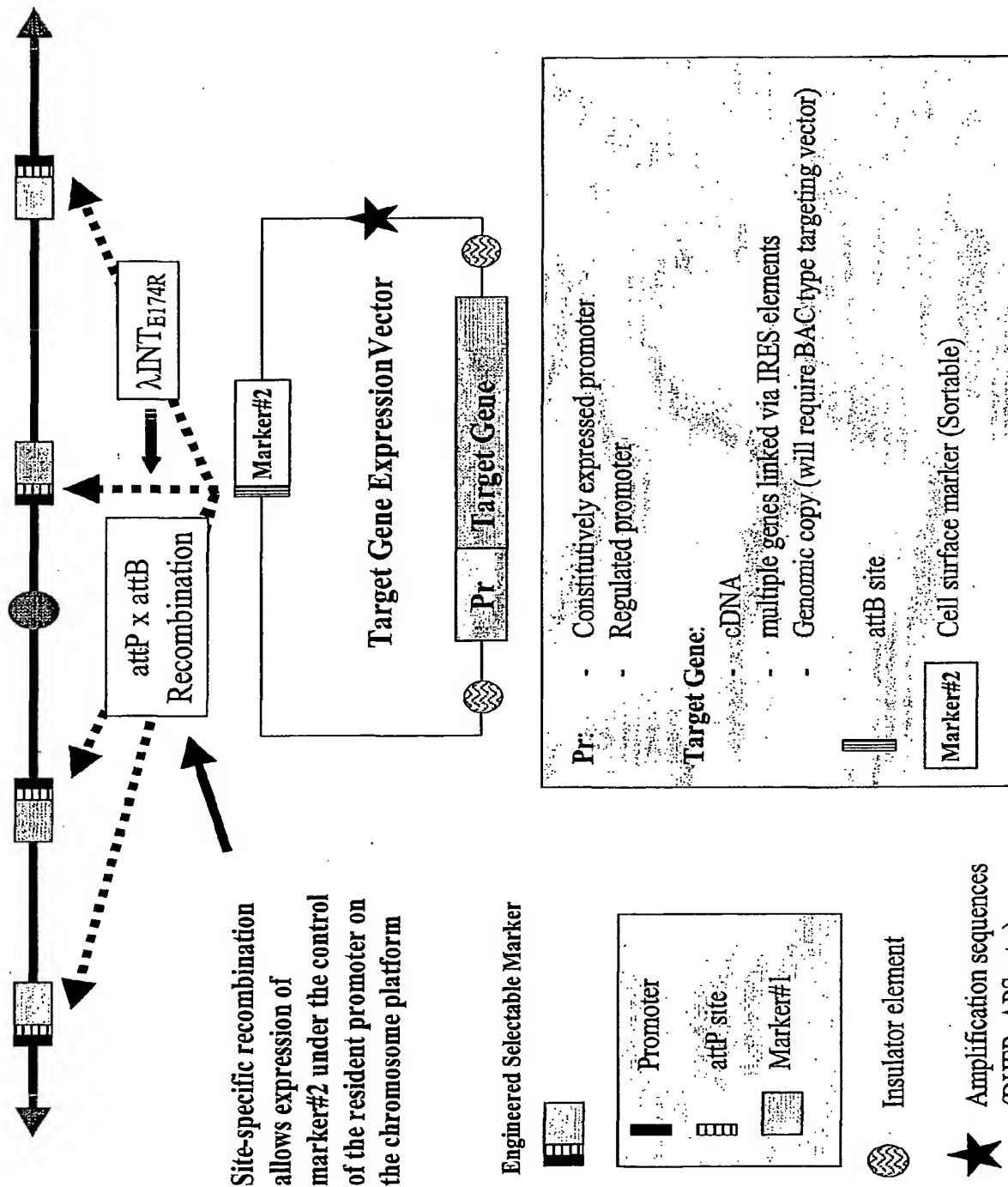




Fig. 10

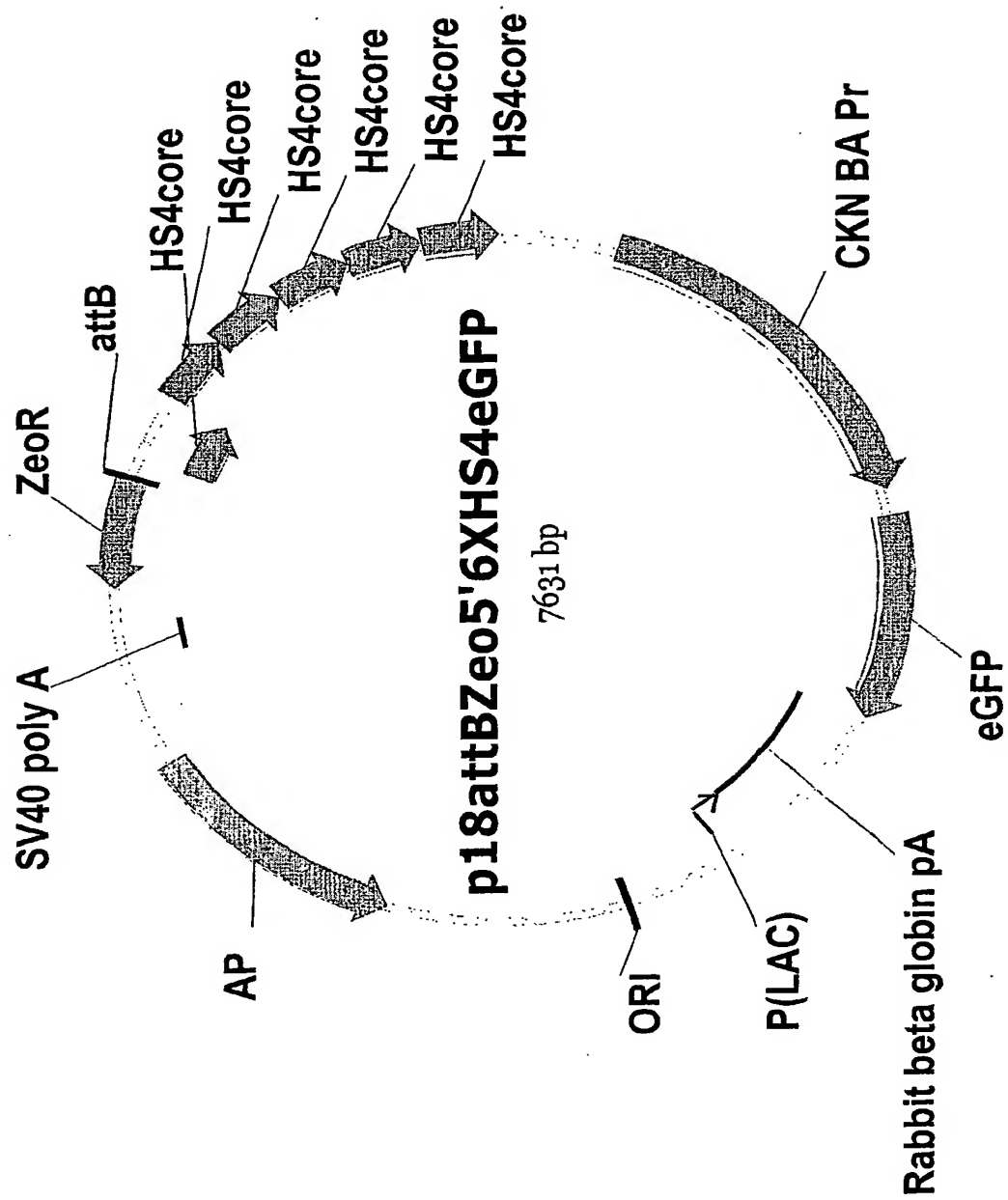
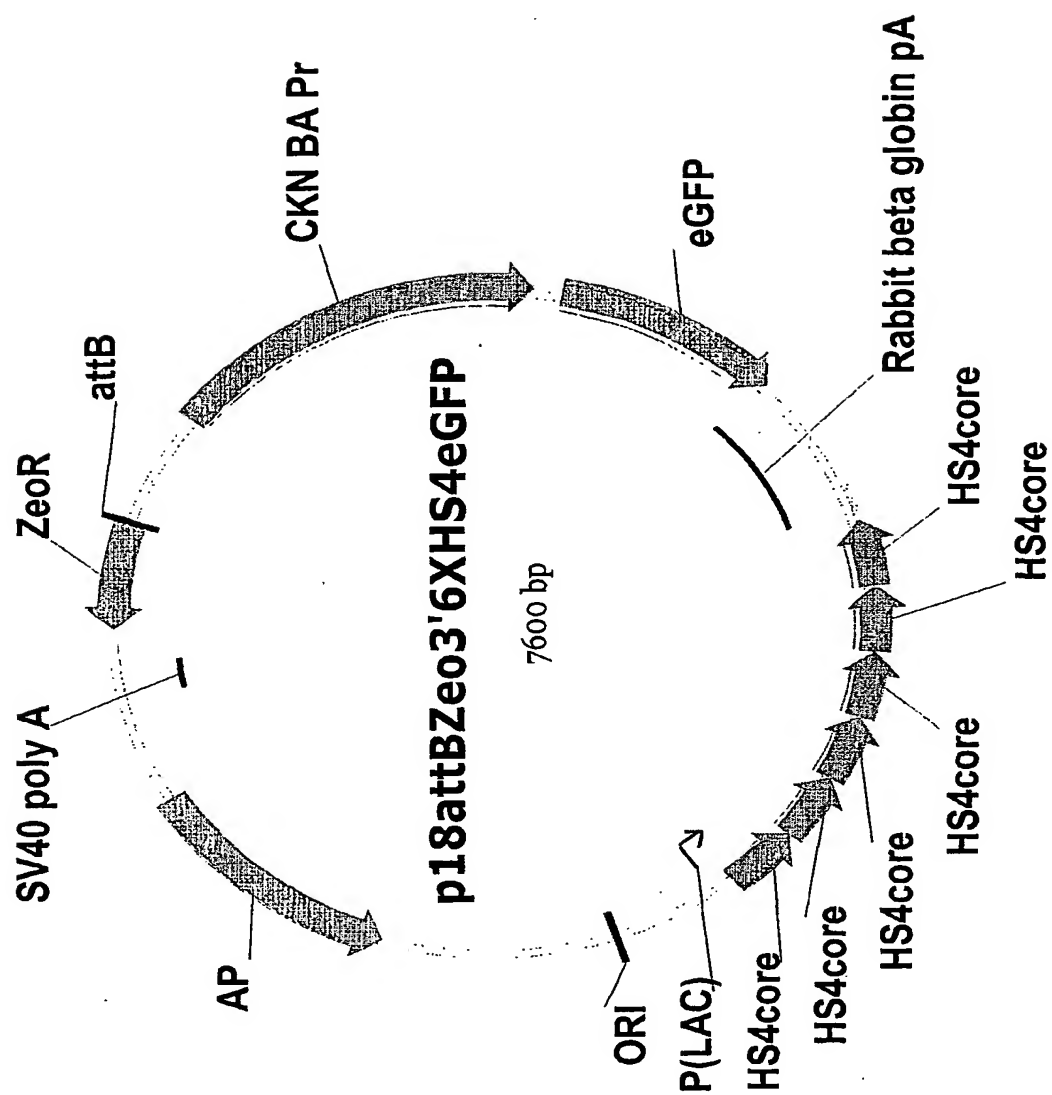


Fig. 11



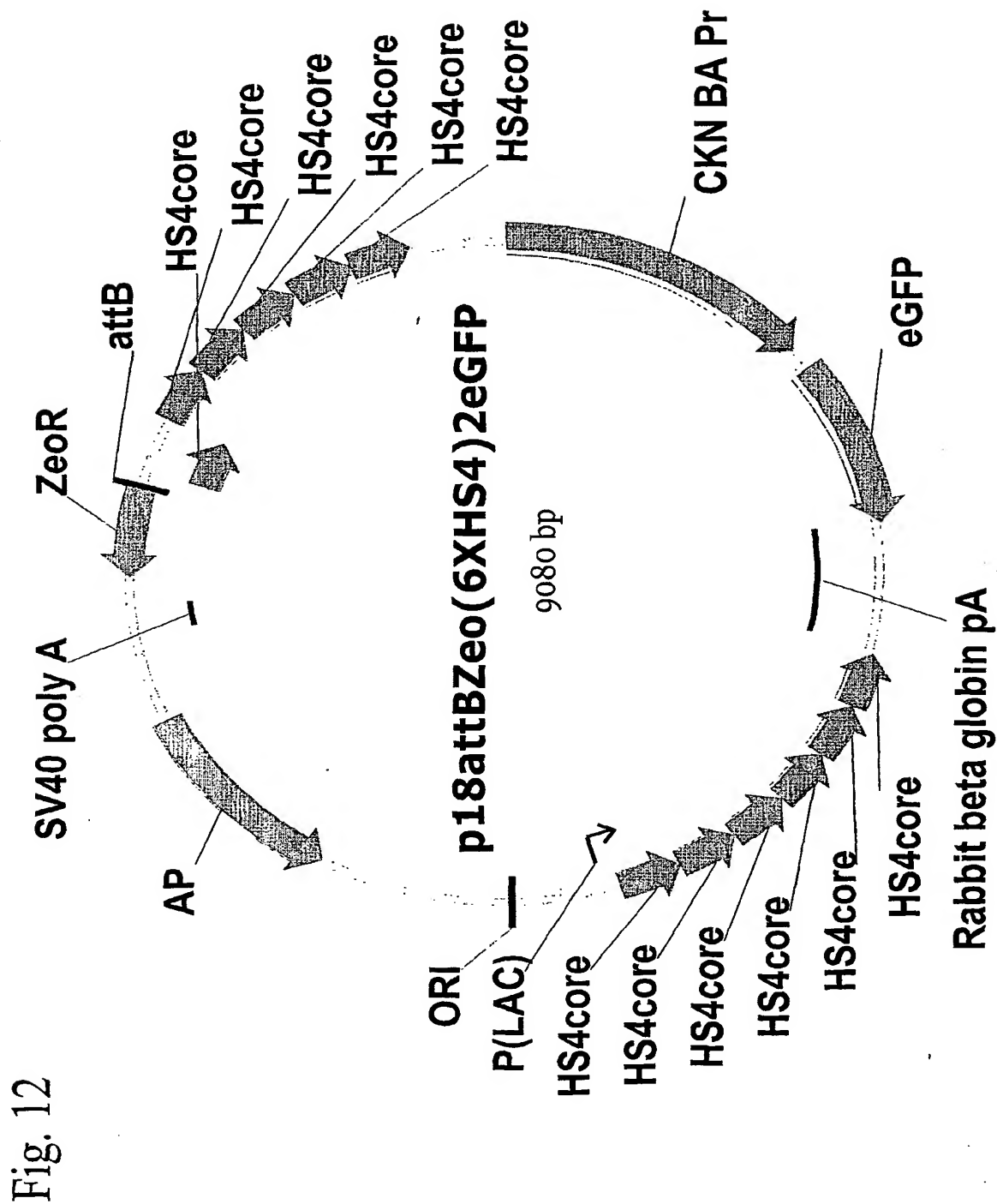
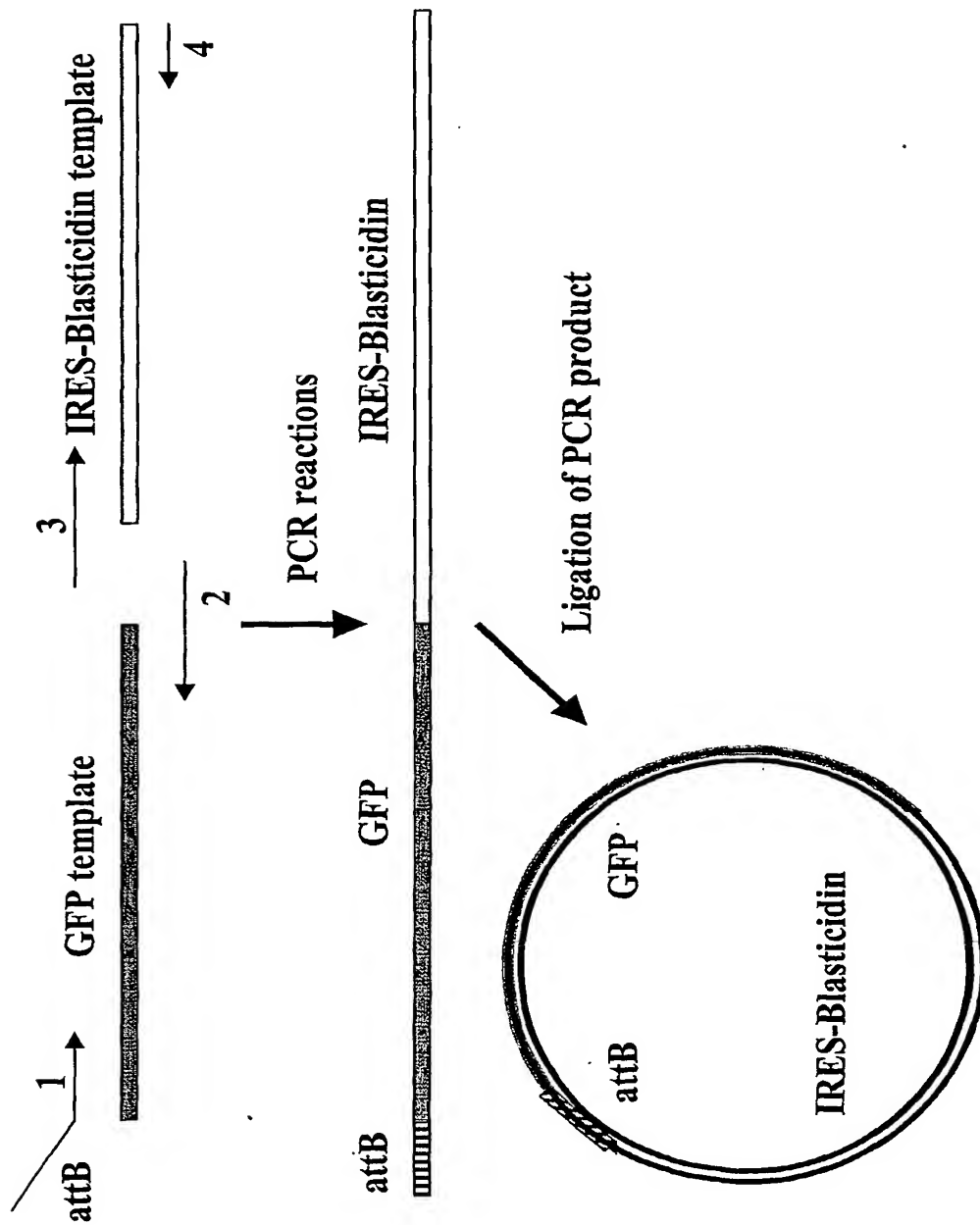


Fig. 13



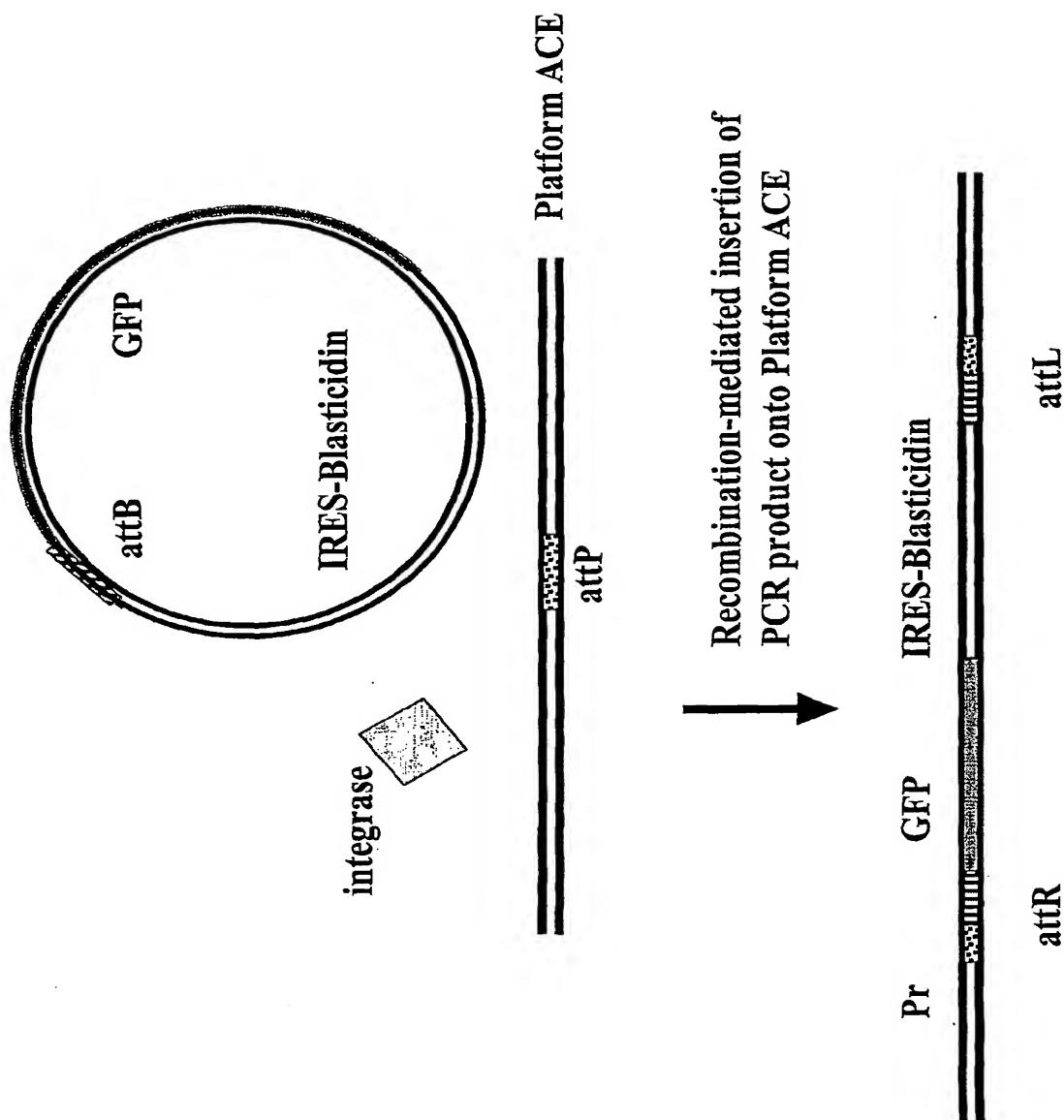


Fig. 14

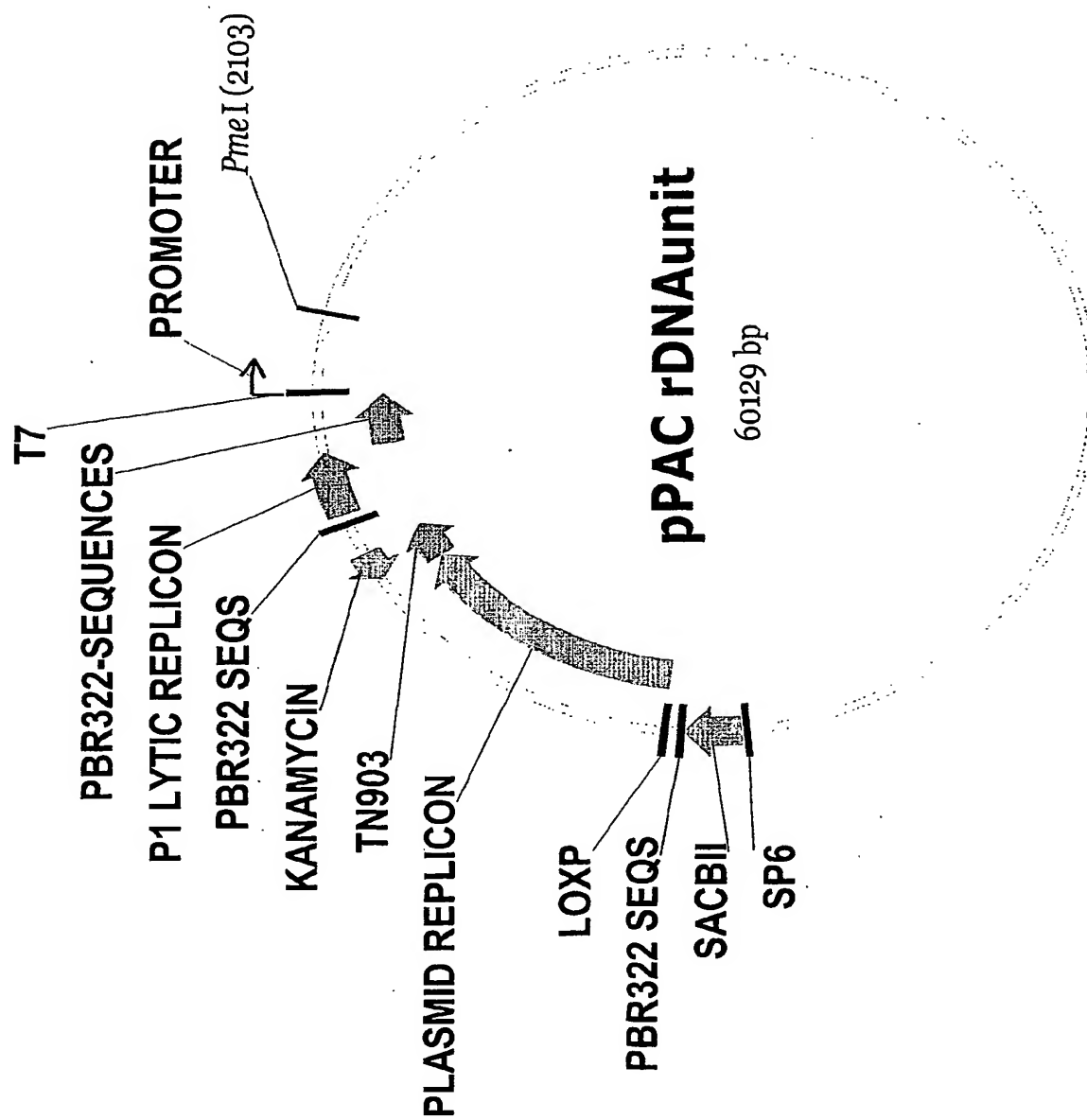


Fig. 15

-1-

## SEQUENCE LISTING

<110> CHROMOS MOLECULAR SYSTEMS, INC.  
Perkins, Edward  
Perez, Carl  
Lindenbaum, Michael  
Greene, Amy  
Leung, Josephine  
Fleming, Elena  
Stewart, Sandra  
Shellard, Joan

<120> CHROMOSOME-BASED PLATFORMS

<130> 24601-420PC

<140> Not Yet Assigned  
<141> Herewith

<150> 60/294,758  
<151> 2001-05-30

<150> 60/366,891  
<151> 2002-03-21

<160> 129

<170> FastSEQ for Windows Version 4.0

<210> 1  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer: attPUP

<400> 1  
ccttgcgcta atgctctgtt acagg 25

<210> 2  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer: attPDWN

<400> 2  
cagaggcagg gaggggaca aaattg 26

<210> 3  
<211> 35  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer: Lamint 1

<400> 3  
ttcgaattca tgggaagaag gcgaagtcac gagcg 35

<210> 4  
<211> 34  
<212> DNA  
<213> Artificial Sequence

-2-

<220>  
<223> Primer: Lamint 2

<400> 4  
ttcgaattct tatttgattt caattttgtc ccac 34

<210> 5  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 5  
cggacaatgc ggttgtgcgt 20

<210> 6  
<211> 46  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> primer

<400> 6  
cgcgcagcaa aatctagagt aaggagatca agacttacgg ctgacg 46

<210> 7  
<211> 46  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> LambdaINTER174rev

<400> 7  
cgtcagccgt aagtcttgat ctcttactc tagattttgc tgcgcg 46

<210> 8  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attB1

<400> 8  
tgaagcctgc ttttttatac taacttgagc gaa 33

<210> 9  
<211> 33  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attB2

<400> 9  
ttcgctcaag ttagtataaa aaagcaggct tca 33

<210> 10  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>



-3-

<223> Primer: attPdown2

<400> 10  
tcttctcggg cataagtcgg acacc 25

<210> 11  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer:CMVen

<400> 11  
ctcacgggga tttccaagtc tccac 25

<210> 12  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer:attPdown

<400> 12  
cagaggcagg gagtgggaca aaattg 26

<210> 13  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer:CMVEN2

<400> 13  
caactccgcc ccattgacgc aaatg 25

<210> 14  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer:L1

<400> 14  
agtatcgccg aacgattagc ttttca 26

<210> 15  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer:F1 rev

<400> 15  
gccgatttcg gcctattggt taaa 24

<210> 16  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer:RED

-4-

<400> 16  
ccgccgacat ccccgactac aagaa

25

<210> 17  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer:L2rev

<400> 17  
ttccttcgaa ggggatccgc ctacc

25

<210> 18  
<211> 22118  
<212> DNA  
<213> Mus musculus

<300>  
<308> GenBank X82564  
<309> 1996-04-09

<400> 18  
gaattcccct atcccctaate cagattgggtg gaataacttg gtatagatgt ttgtgcatta 60  
aaaaccctgt aggatcttca ctctagggtca ctgttcagca ctggaaacctg aattgtggcc 120  
ctgagtata ggtcctggga catatgcagt tctgcacaga cagacagaca gacagacaga 180  
cagacagaca gacagacgtt acaaaacaaac acgttgagcc gtgtgccaac acacacacaa 240  
acaccactct ggccataaatt attgaggacg ttgatttatt attctgtgtt tgtgagtctg 300  
tctgtctgtc tgtctgtctg tctgtctgtc tatcaaacca aaagaaacca aacaattatg 360  
cctgcctgcc tgcctgctg cctacacaga gaaatgattt cttcaatcaa tctaaaacga 420  
cctcctaagt ttgccttttt tctctttctt tatcttttct ttttttcttt tcttcttctt 480  
tcttcttctt tcttcttctt ctttctttct ttctttcttt ttctttcttt cttactttct 540  
ttctttcttt cttacatttta ttcttttcat acatagtttc ttagtgtaag catccctgac 600  
tgtcttgaag acactttgta ggcctcaatc ctgtaagagc cttoctctgc ttttcaaattg 660  
ctggcatgaa tgttgtacct cactatgacc agcttagtct tcaagtctga gttactggaa 720  
aggagtcca agaagactgg ttatatTTTT catttattat tgcattttta ttaaaattta 780  
atttcaccaa aagaatttag actgaccaat tcagagtctg ccgtttaaaa gcataaggaa 840  
aaagtaggag aaaaacgtga ggctgtctgt ggatggctga ggctgcttta gggagcctcg 900  
tcaccattct gcacttgcaa accgggccac tagaaccggg tgaagggaga aaccaaagcg 960  
acctggaaac aataggtcac atgaaggcca gccacctcca tcttggtgtg cgggagttca 1020  
gttagcagac aagatggctg ccatgcacat gttgtctttc agcttggtga ggtcaaagta 1080  
caaccgagtc acagaacaag gaagtataca cagttagttc caggtcagcc agagtttaca 1140  
cagagaaacc acatcttgaa aaaaaacaaa aaataaatta aataaataata atttaaaaat 1200  
ttaaaaatag ccgggagtg tggcgcatgt ctttaatccc agctctcttc aggcagagat 1260  
gggaggattt ctgagtttga ggccagcctg gtctgcaaag tgagttccag gacagtcagg 1320  
gctatacaga gaaaccctgt cttgaaaact aaactaaatt aaactaaact aaactaaaaa 1380  
aatataaaat aaaaatttta aagaatttta aaaaactaca gaaatcaaac ataagcccc 1440  
gagatggcaa gtaactgcaa tcatagcaga aatattatac acacacacac acacagactc 1500  
tgtcataaaa tccaatgtgc cttcatgatg atcaaatttc gatagtcagt aatactagaa 1560  
gaatcatatg tctgaaaata aaagccagaa ccttttctgc ttttggtttc ttttgcccca 1620  
agatagggtt tctctcagtg tatcctggc atccctgcct ggaacttctt ttgtagggtt 1680  
ggtagcctca aactcagaga ggtcctctct gcctgcctgc ctgcctgcct gcctgcctgc 1740  
ctgcctgcct gcctgcctca cttcttctgc caccacacac accgagtcga acctaggatc 1800  
tttatttctt tctctttctc tcttctttct ttctttcttt ctttctttct ttctttcttt 1860  
ctttctttct ttcttattca attagttttc aatgtaagtg tgtgtttgtg ctctatctgc 1920  
tgcctatagg cctgtattgc aggagagggc aacagaacct aggagaaaac acctgacag 1980  
tcctgagaat aagtgaaaaa acaacaaaaa aaggaaattc taatcacata gaatgtagat 2040  
atatgccgag gctgtcagag tgctttttta ggcttagtgt aagtaatgaa aattgttgtg 2100  
tgtcttttat ccaaacacag aagagaggtg gctcggcctg catgtctgtt gtctgcattg 2160  
agaccaggct gccttgaaac acattaatct gctcgcctaa gcttccctaa tgtgctccac 2220  
aaaggcatgt gccaccactg cccggactga tttcttcttt tttttttttt tggaaaatac 2280  
ctttctttct ttttctctct ctcttttctt cttcttctct ttctttctat tctttttttc 2340  
tttctttttt tttttttttt aatttgctta aggttaaaagg tgtgtccac 2400  
aattgcctca gctctgctct aattctcttt aaaaaaaaac aaacaaaaaa aaaacccaaa 2460  
cagtatgtat gtatgtatat ttagaagaaa tactaatcca ttaataactc ttttttctta 2520  
aaattcatgt cattcttggt ccacaaagtg agttccagga cttaccagag aaaccctgtg 2580

ttcaaatttc	tgtgttcaag	gtcaccctgg	cttacaaagt	gagttccaag	tccgataggg	2640
ctacacagaa	aaaccataac	tcagaaaaaa	aaaaagtcc	aaacacacac	acacacacac	2700
acacacacac	acacacacac	acacacacac	acacacacag	cgcgcccgcg	cgatgagggg	2760
aagtctgtcc	taaaataaat	atTTTTctgg	ccaaagtga	agcaaatcac	tatgaagagg	2820
tactcctaga	aaaaataaat	acaaacgggc	TTTTtaatca	ttccagcact	gttttaattt	2880
aactctgaat	ttagtcttgg	aaaagggggc	gggtgtgggt	gagtgagggc	gagcgagcag	2940
acgggcgggc	gggcggggtga	gtggccggcg	gcgggtggcag	cgagcaccag	aaaacaacaa	3000
acccccagcg	gtagagtggt	ttaaaaatga	gacctaaatg	tgggtggaacg	gaggtcgccg	3060
ccaccctcct	cttccactgc	ttagatgctc	ccttccccct	actgtgctcc	cttccccctaa	3120
ctgtgcctaa	ctgtgcctgt	tccctcaccc	cgctgattcg	ccagcgacgt	actttgactt	3180
caagaacgat	tttgccctgtt	ttcaccgctc	cctgtcatac	tttcgttttt	gggtgccccg	3240
gtctagcccg	ttcgctatgt	tcgggcgggg	cgatggggac	cgtttgtgcc	actcgggaga	3300
agtggtgggt	gggtacgctg	ctccgtcgtg	cgctgctgag	tgccggaacc	tgagctcggg	3360
agaccctccg	gagagacaga	atgagtgaat	gaatgtggcg	gcgcgtgacg	gatctgtatt	3420
ggtttgtatg	tttgatcgag	accattgtcg	ggcgacacct	agtggtgaca	agtttcggga	3480
acgctccagg	cctctcaggt	tggtgacaca	ggagagggaa	gtgcctgtgg	tgaggcgacc	3540
aggggtgacag	gaggccgggc	aagcaggcgg	gagcgtctcg	gagatgggtg	cgtgtttaag	3600
gacggtctct	aacaaggagg	tcgtacaggg	agatggcctc	agcagaccga	gttgctgtac	3660
gcccttttgg	gaaaaatgct	aggggttgggt	gcaacgttac	taggtcgacc	agaaggctta	3720
agtcctaccc	ccccccccct	tttttttttt	tttccctccag	aagccctctc	ttgtccccgt	3780
caccgggggc	accgtacatc	tgaggccgag	aggacggcat	gggccccggt	tcacaagccg	3840
tgtggctcgg	ccagctggcg	cttcgggtct	tttttttttt	tttttttttt	ttttcctcca	3900
gaagccttgt	ctgtcgctgt	caccgggggc	gctgtacttc	tgaggccgag	aggacgcgat	3960
gggccccggc	ttccaagccg	gtgtggctcg	gccagctgga	gcttcgggtc	tttttttttt	4020
tttttttttt	tttttttctc	cagaagcctt	gtctgtcgct	gtcaccgggg	gcgctgtact	4080
tctgaggccg	agaggacggc	atgggtcggc	ttccaagccg	atgtggccgg	gccagctgga	4140
gcttcgggtt	tttttttttt	ctccagaagc	cctctcttgt	ccccgtcacc	ggggcgcgct	4200
tacttctgag	gccgagagga	cgtgatgggc	ccgggttcca	ggcggatgtc	ggccgggtcag	4260
ctggagcttt	ggatcctttt	tttttttttt	cctccagaag	ccctctcttg	tcctccgtcac	4320
cgggggcacc	ttacatctga	gggcgagagg	acgtgatggg	tccggcttcc	aagccgatgt	4380
gctggggcca	gctggagctt	cggttttttt	tttttctctc	cagaagccct	ctcttgtccc	4440
cgtcacccgg	ggcgtgttac	ttctgaggcc	gagaggacgt	gatgggcccc	ggttccaggc	4500
ggatgtcgcc	cggtcagctg	gagctttgga	tcattttttt	ttttccctcc	agaagccctc	4560
tctgtcccc	gtcacccggg	gcaccgtata	tctgaggccg	agaggacacg	atgggccccgt	4620
cttccaagcc	gatgtggccc	ggccagctgg	agcttcgggt	cttttttttt	ttttttccctc	4680
cagaagcctt	gtctgtcgct	gtcacccggg	gcgctgtact	tctgaggccg	agaggacggc	4740
atgggccccg	gttccaagcc	gggtgtggct	ggccagctgg	agcttcgggt	cttttttttt	4800
tttttttttt	ttcctccaga	aaccttgtct	gtcgctgtca	cccgggggcg	ttgtacttct	4860
gatgccgaga	ggacgcgatg	ggcccgctct	ccaggccgat	gtggccccgt	cagctggagc	4920
tttggatctt	tttttttttt	tttccctcca	gaagccctct	cttgtccccg	tcaccggggg	4980
caccttacat	ctgaggccct	gaggacacga	tgggccccgg	ttccaggccg	atgtggcccc	5040
gtcagctgga	gctttggatc	tttttttttt	ttttcttcca	gaagccctct	tgtccccgtc	5100
accggtggca	ctgtacatct	gaggccggag	ggacattatg	ggccccggct	ccaatccgat	5160
gtggccccgt	cagctggagc	tttggatctt	atTTTTtttt	taattttttc	ttccagaagc	5220
cctcttgtcc	ctgtcacccg	tggcacggta	catctgaggc	cgagaggaca	ttatgggccc	5280
ggcttccagg	ccgatgtggc	ccggtcagct	ggagctttgg	atcttttttt	ttttttttct	5340
tttttccctc	agaagccctc	tctgtccctg	tcaccggggg	ccctgtacgt	ctgaggccga	5400
gggaaagcta	tgggcgcggt	tttctttcat	tgacctgtcg	gtcttatcag	ttctccgggt	5460
gttcagggtc	gaccagttgt	tcctttgagg	tcgggttctt	ttcgttatgg	ggtcattttt	5520
gggccacctc	cccaggatag	acttccaggc	gtcgltgtct	gcctgtcact	ttcctccctg	5580
tctcttttat	gcttgtgatc	ttttctatct	gttccctattg	gacctggaga	taggtactga	5640
cacgctgtcc	tttccctatt	aacactaaag	gacactataa	agagaccctt	tcgatttaag	5700
gctgttttgc	ttgtccagcc	tattcttttt	actggcttgg	gtctgtcgcg	gtgcctgaag	5760
ctgtccccga	gccacgcttc	ctgctttccc	gggcttgcgt	cttgctgtgt	cttgctgtgg	5820
gcagcttgtg	acaactgggc	gctgtgactt	tgctgcgtgt	cagacgtttt	tcctcgattt	5880
cccagggtgt	cgttgtcaca	cctgtccccg	ttggaatggg	ggagccagct	gtgggtgagg	5940
gccaccttat	ttcggtcac	tttttttttt	tttttttctc	ttggagctcc	gaacctccgc	6000
tcttttctct	tcctgggtctt	tttttttttt	gcctcccgag	tgcatttctt	tttgtttttt	6060
ttcttttttt	tttttttttt	ttggggaggt	ggagagtccc	gagtaactca	ctcctgtctg	6120
tggtgtccaa	gtgttcatgc	cacgtgcctc	ccgagtgcac	ttttttttgt	ggcagtcgct	6180
cgttgtgttc	tgctgttctg	tgtctgcccg	tatcagtaac	tgtcttgccc	cgctgtgaag	6240
acattcctat	ctcgcttgtt	tctcccgatt	gcgctcgctt	gctcactctt	agatcgatgt	6300
ggtgtccggg	agttctcttc	gggcccgggc	caagccgcgc	caggcgaggg	acggacattc	6360
atggcgaatg	gcggccgctc	ttctcgctct	gccagcgggc	cctcgctctc	ccaccccatc	6420
cgtctgcccg	tggtgtgtgg	aaggcagggg	tgcggctctc	cgccccgacg	ctgcccccg	6480
cgcacttttc	tcagtgggtc	gcgtggctct	tgtggatgtg	tgaggcgccc	ggttgtgccc	6540
tcacgtgttt	cactttgggc	gtgtctcgct	tgaccatggt	cccagagtcg	gtggatgtgg	6600

ccgggtggcgt	tgcataccct	tcccgtctgg	tgtgtgcacg	cgctgtttct	tgtaagcgtc	6660
gaggtgctcc	tggagcgttc	caggtttgtc	tcctaggtgc	ctgcttctga	gctgggtggg	6720
gcgctcccca	ttccctgggtg	tgcctccggg	gctccgtctg	gctgtgtgcc	ttcccgtttg	6780
tgtctgagaa	gcccgtgaga	gggggggtcga	ggagagaagg	aggggcaaga	cccccttct	6840
tcgtcgggtg	aggcgccac	cccgcgacta	gtacgcctgt	gcgtagggt	gggtgctgagc	6900
ggtcgcggct	ggggttggaa	agtttctcga	gagactcatt	gctttcccg	ggggagcttt	6960
gagaggcctg	gctttcgggg	gggaccgggt	gcagggtctc	ccctgtccgc	ggatgctcag	7020
aatgcccttg	gaagagaacc	ttcctgttgc	cgcagacccc	cccgcgcgg	cgcccgctg	7080
ttggctctct	ggtttccctg	tgtgtcgtc	gcatgcatcc	tctctcggtg	gcccgggctc	7140
gtcgggggtt	tgggtccgtc	ccgcccctcag	tgagaaagtt	tccttctcta	gctatcttcc	7200
ggaaaggggtg	cgggcttctt	acgggtctcga	gggggtctctc	ccgaatgggtc	ccctggagg	7260
ctcgcccccct	gacgcctcc	cgcgcgcgca	gcgtttgtctc	tctcgtctac	cgcgcccg	7320
ggcctccccg	ctccgagttc	ggggagggtat	cacgcggggc	agagcctgtc	tgtcgtcctg	7380
ccgttgctgc	ggagcatgtg	gctcggcttg	tgtgggtggg	ggctggggag	agggtccgt	7440
gcacaccccc	gcgtgcgcgt	actttctctc	cctcctgagg	gccgcggtg	ggacggggg	7500
tggttaggcg	acgggtgggt	cccggttccc	caccgcctct	cccggtgcctc	acccgtgcct	7560
tcctgcgcgt	gcgtccctct	cgctcgcgtc	cacgactttg	gccgctcccg	cgacggcg	7620
ctgcgcgcgc	cgtgggtcgt	gctgtgtgct	tctcgggtg	tgtgggtgtg	tcgcctcgcc	7680
cccccttccc	cgcggcagcg	ttcccacggc	tggcgaaatc	gcgggagtc	tccttcccc	7740
cctcgggggt	gagaggggtc	gtgtctggcg	ttgattgatc	tcgctctcgg	ggacgggacc	7800
gttctgtggg	agaacggctg	ttggccgcg	ccggcgcgac	gtcggacgtg	gggaccacct	7860
gccgctcggg	ggctcttcgtc	ggttaggcac	ggtgtgtcgg	catcggtctc	tctctcgtgt	7920
cgggtgcgcc	tcctcgggtc	cccggggggc	cgtcgtgttt	cgggtcgggt	cggcgctgca	7980
gggtgtgtgg	gactgctcag	gggagtgggt	cagtgtgatt	cccgcgggtt	ttgcctcgcg	8040
tgccctgacc	ggctccgacgc	ccgagcgggtc	tctcgggtccc	ttgtgaggac	ccccttccgg	8100
gaggggcccc	tttcggccgc	cettgcgcgtc	gtcgcggggc	ctcgttctgc	tgtgtcgttc	8160
ccccctcccc	gctcgccgca	gcccgtcttt	tttctctct	ccccccctct	cctctgaactg	8220
acccgtggcc	gtgctgtcgg	accccccgca	tgggggcggc	cgggcacgta	cgcgctccggg	8280
cgggtaccgg	ggctcttgggg	gggggcccag	gggtaagaaa	gtcggctcgg	cgggcgggag	8340
gagctgtggg	ttggaggcg	tcccggcccc	cggcgcgtgg	cgggtgtctt	cgcggtcttg	8400
gagagggtcg	cgtgcgaggg	gaaaagggtg	ccccgcgagg	gcaaaggga	agaggctagc	8460
agtgtgtcatt	gtccccgacg	tgtgtgtgtg	tgttggccga	ggtgcgtctg	gggggctcgt	8520
ccggccctgt	gctccgtcgg	gaaggcgcgt	gttggggcct	gccggagtgc	cgaggtgggt	8580
accctggcgg	tgggattaac	cccgcgcgcg	tgtcccgggtg	tggcgggtggg	ggctccgggtc	8640
gatgtctacc	tcctctctccc	cgaggtctca	ggccttctcc	gcgcggggctc	tcggccctcc	8700
cctcgttccc	ccctctcgcg	gggttcaagt	cgctctcga	cctccccctcc	tcctccttcc	8760
catctctcgc	gcaatggcgc	cgcccagatt	cacgggtgggt	tcgtcctccg	cctccgcttc	8820
tcgcgcgggg	ctggccgctg	tccgggtctct	cctgcccagac	ccccgttggc	gtgggtcttt	8880
ctcgcgggct	ctcgcgactc	ctggcttcgc	cggagggttc	agggggcttc	cgggttcccc	8940
gacgttgccg	ctcgcgtcgt	tgtgtctggg	ggggggccgc	tgcggcctcc	gcccgcctcg	9000
gagccctcgc	cgcacccgcc	ggtgtgcggg	ttcgcgcgcg	ggtcagttgg	gccctggcgt	9060
tgtgtcgcgt	cgggagcgtg	tccgcctcgc	ggcggctaga	cgcgggtgtc	gcccgggtcc	9120
gacgggtggc	ctatccaggg	ctcgcgcccg	ccgacccccg	cctgcccgtc	ccggtgggtg	9180
tcgttgggtg	ggggagtga	tgggtctacc	ggtcattccc	tcccgcgtgg	tttgactgtc	9240
tcgcgggtgt	cgcgcttctc	tttccgccaa	ccccacgcgc	aaccaccac	cctgtctccc	9300
cggcccggtg	cggtcgacgt	tccggtctct	ccgatgccga	gggggttcggg	atttgtgcgg	9360
gggacggagg	ggagagcggg	taagagaggt	gtcggagagc	tgtcccgggg	cgacgctcgg	9420
ggtggctttg	ccgcgtgcgt	gtgctcgcgg	acgggttttg	tcggaccccc	acggggtcgg	9480
tcgggcgcga	tgcactctcc	cgttccgcgc	gagcgcgcgc	ccgggtcacc	cccggtttgt	9540
cctccgcgca	ggctctccgc	cgcgcgcgc	tcctcctcct	ctctcgcgct	ctctgtcccc	9600
cctggtcctg	tcccaccccc	gacgctccgc	tcgcgcttcc	ttacctgggt	gatcctgcca	9660
ggtagcata	gcttgtctca	aagattaagc	catgcatgtc	taagtacgca	cggccgggtac	9720
agtgaactg	cgaatggctc	attaaatcag	ttatggttcc	tttgggtcgt	cgctcctctc	9780
ctacttggat	aactgtggtg	attctagagc	taatacatgc	cgacggggcg	tgacccccct	9840
tcccgggggg	ggatgcgtgc	atttatcaga	tcaaaaccac	cccgggtgagc	tcctccccgg	9900
ctccggccgg	gggtcggg	ccggcggctt	ggtgactcta	gataacctcg	ggccgatcgc	9960
acgccccccg	tggcggcgac	gacccattcg	accgtctgcc	ctatcaactt	tcgatggtag	10020
tcgcgctgcc	taccattggtg	accacgggtg	acggggaatc	agggttcgat	tcggagagg	10080
gagcctgaga	aacggctacc	acatccaagg	aaggcagcag	gcgcgcaaat	taccactcc	10140
gcacccgggg	aggtagtgac	gaaaaataac	aatacaggac	tctttcgagg	ccctgtaatt	10200
ggaatgagtc	cactttaaat	cctttaacga	ggatccattg	gagggcaagt	ctgggtccag	10260
cagccgcgggt	aattccagct	ccaatagcgt	atattaaagt	tgctgcagtt	aaaaagctcg	10320
tagggtgatc	ttgggagcgg	gcgggcgggtc	cgccgcgagg	cgagtcaccg	cccgcccccg	10380
cccttgcctt	cctcgatgct	cctcgatgct	cttagctgag	tgtcccgcgg	ggcccgaagc	10440
gtttactttg	aaaaaattag	agtgttcaaa	gcaggcccga	gccgcctgga	taccgcagct	10500
aggaataatg	gaataggacc	gcgggttctat	tttgttggtt	ttcggaactg	aggccatgat	10560
taagagggac	ggccgggggc	attcgtattg	cgccgctaga	ggtgaaattc	ttggaccggc	10620

gcaagacgga	ccagagcgaa	agcatttgcc	aagaatgttt	tcattaatca	agaacgaaa	10680
tcggaggttc	gaagacgatc	agataccgtc	gtagttccga	ccataaacga	tgccgactgg	10740
cgatgcggcg	gcgttatcc	catgacccgc	cgggcagctt	ccgggaaacc	aaagtctttg	10800
gggtccgggg	ggagtatgg	tgcaaaagctg	aaacttaaa	gaattgacgg	aagggcacca	10860
ccaggagtg	gcctgcggct	taatttgact	caacacggga	aacctcacc	ggcccgga	10920
cggacaggt	tgacagattg	atagctcttt	ctcgattccg	tggggtggg	tgcatggccg	10980
ttcttagttg	gtggagcgat	ttgtctgggt	aattccgata	acgaacgaga	ctctggcatg	11040
ctaactagtt	acgcgacccc	cgagcggtcg	gcgtccccc	acttcttaga	gggacaagtg	11100
gcgttcagcc	acccgagatt	gagcaataac	aggtctgtga	tgcccttaga	tgtccggggc	11160
tgacgcgcg	ctacactgac	tggctcagcg	tgtgcctacc	ctgcgcggc	aggcgcggt	11220
aaccggtga	accccatctg	tgatggggat	cggggattgc	aattattccc	catgaacgag	11280
gaattccag	taagtccggg	tcataagctt	gcgttgatta	agtcctggc	ctttgtacac	11340
accgcccgtc	gctactaccg	attggatgg	ttagtgaggc	cctcggatcg	gccccgcgg	11400
ggtcggccca	cggccctggc	ggagcgctga	gaagacggc	gaacttgact	atctagagga	11460
agtaaaagtc	gtcaacaagg	ttccgtagg	gaacctggc	aaggatcatt	aaacgggaga	11520
ctgtggagga	gcggcgcggt	ggcccgctct	ccccgtcttg	tgtgtgtcct	cgccgggagg	11580
cgctgcgtc	ccgggtcccc	tcgcccgcgt	gtggagcgag	gtgtctggag	tgaggtgaga	11640
gaaggggttg	gtggggtcgg	tcgggtccg	ctggggaccg	cctccgattt	ccccccccc	11700
ttccctctcc	ctcgctccgg	tcgtacctcg	ccaccctacc	gcggcgcgcg	ctgctcgcg	11760
gcgtcttgcc	tccttcccg	ccggctcttc	cgctgtctacg	aggggcggta	cgctggttacg	11820
ggtttttgac	ccgtcccggg	ggcgttcgg	cgctggggcg	cgcgctttgc	tctcccgga	11880
cccatccccc	ccgcggctct	ggcttttcta	cgctggctgg	ggcggttgct	gcgtgtgggg	11940
ggatgtaggt	ctcgcggtgtg	ggctcgcccg	tccegatgcc	acgcttttct	ggcctcgctg	12000
gtcctcccc	gtcctgtccc	gggtacctag	ctgtcgctt	ccggcgcgga	ggtttaagga	12060
ccccgggggg	gtcgccctgc	cgcccccagg	gtcggggggc	ggtggggccc	gtagggaggt	12120
cggtcggtcg	ggcggtctct	cctcagactc	catgaccctc	ctccccccgc	tgccgcggt	12180
cccagggcgg	cggtcggtgtg	gggggggtgg	tgtctggagc	cccctcgggc	gcccgtgggg	12240
cccagaccgg	gcccgcggct	tgcccgat	ccgcgggtcg	gtcctgtcgg	tgccgggtcgt	12300
gggttcccg	gtcggttccc	tgtttttccg	ctcccgacc	tttttttttc	ctcccccca	12360
caegtgtctc	gtttcggtcc	tgctggccgg	cctgaggcta	cccctcggtc	catctgttct	12420
ctctctctct	cggggagagg	agggcggtgg	tcgttggggg	actgtgccc	cgctcagcacc	12480
cgtgagttcg	ctcacaccgg	aaataaccgat	acgactctta	gcgggtggatc	actcggctcg	12540
tgctgcgatg	aagaacgcag	ctagctgcga	gaattaatgt	gaattgcagg	acacattgat	12600
catcgacact	tcgaacgcac	ttgcggccccc	gggttctctc	cggggctacg	cctgtctgag	12660
cgtcggttga	cgatcaatcg	cgtcacccgc	tgccgtgggt	gctgcgcggc	tgggagtttg	12720
ctcgcagggg	caacccccca	acccgggtcg	ggccctccgt	ctcccgaggt	tcagacgtgt	12780
gggcgggtgt	cgggtgtggcg	cgcgcgcccg	cgctcgggag	cctgggtctcc	cccgcgcac	12840
cgcgctcgcg	gcttcttccc	gctccgccc	tccegcctc	gcccgtgcac	cccggctcctg	12900
gcctcgcgct	ggcgccctccc	ggaccgctgc	ctcaccagtc	tttctcggtc	ccgtgcccgc	12960
tgggaaccca	cgcgccccc	gtggcgcccg	gggggtggcg	cgctccgcac	tgctctgggtc	13020
gaggttggcg	gttgagggtg	tgctgtcgcc	gaggtgggtg	tcgggtccc	gcggccgcgg	13080
gggttgtcggg	gtggcggtcg	acgagggccg	gtcggtcgcc	tgccgtgggt	gtctgtgtgt	13140
gtttgggtct	tgcgctgggg	gagggcggtg	cgaccgctcg	cgggggttggc	gcgggtcgcc	13200
ggcgccgcgc	accctccggc	ttgtgtggag	ggagagcgag	ggcgagaaac	gagagaggtg	13260
gtatccccgg	tgccgttgcc	agggaggggt	tggcgctccc	cgctccgtccg	tcctctccctc	13320
cctcggtggg	cgccttcggc	ccgcacgcgg	ccgctagggg	cggtcggggc	ccgtggcccc	13380
cgtggctctt	cttcgtctcc	gcttctcctt	cacccgggcg	gtaccgctc	cgcgcccggc	13440
ccgcgggacg	ccgcggcgct	cgtgcgcgga	tgcgagtcac	ccccgggtgt	tgccaggttcg	13500
gggagggaga	gggcctcgct	gaccggttgc	gtcccggctt	ccctggggggg	gaccggcggt	13560
ctgtgggctg	tgcttcccgg	gggttgctgt	tgagtaagat	cctccacccc	cgccgcccctc	13620
ccctcccggc	ggcctctcgg	ggacccccctg	agacggttcg	ccggctcgtc	ctcccgtgce	13680
gccgggtgcc	gtctctttcc	cgcccgccctc	ctcgctctct	tcttcccgcg	gctggggcg	13740
tgccccccct	ttctgaccgc	gacctcagat	cagacgtggc	gaccgctga	atttaagcat	13800
attagtacgc	ggaggaaaag	aaactaacca	ggattccctc	agtaacggcg	agtgaacagg	13860
gaagagccca	gcgcggaatc	cccgcgcgc	gtcgcggcgt	gggaaatgtg	gcgtacggaa	13920
gaccactcc	ccggcgccgc	tcgtgggggg	cccaagtcct	tctgatcgag	gcccagccc	13980
tggacggtgt	gaggccgggt	gcggcccccg	cgcgcggggc	tcgggtcttc	ccggagtcgg	14040
gttgcttggg	aatgcagccc	aaagcgggtg	gtaaaactca	tctaaaggcta	aaacgggca	14100
cgagaccgat	agtcaacaag	taccgtaagg	gaaagttgaa	aagaactttg	aagagagagt	14160
tcaagagggc	gtgaaacggt	taagaggtaa	acgggtgggg	tcgcgcgagt	ccgcccggag	14220
gattcaaccc	ggcggcgcgc	gtccggcggt	cccggttgg	ccggcggtat	ctttcccgct	14280
ccccgttcc	cccgacccct	ccacccgcgc	gtcgttcccc	tcttccctccc	cgctccggc	14340
gctccggcg	gcgggcgcgg	ggggtgggtg	gggtggggcg	cgcgggcggg	gcccgggggtg	14400
gggtcggcg	gggacggcc	ccggccggcg	accggccgcc	gcccggcgca	cttccacgt	14460
ggcggtgcgc	cgcgaccggc	tcggggacgg	ccgggaaggc	ccggtgggga	aggtggctcg	14520
ggggggggcg	cgcgtctcag	ggcgcgccga	accacctcac	cccgaggtgt	acagccctcc	14580
ggccgcgctt	tcgcccgaatc	ccggggccga	ggaagccaga	taccgctcgc	cgcgctctcc	14640

ctctccccc	gtccgcctcc	cgggcgggcg	tgggggtggg	ggccggggcg	ccccccac	14700
ggcgcgaccg	ctctccacc	ccctcccgtc	gcctctctcg	gggcccgggtg	ggggcgggg	14760
cgactgtcc	ccagtgcgcc	ccggcggtcg	tcgcgccgtc	gggtcccggg	gggaccgtcg	14820
gtcacgcgtc	tcccgcagaa	gccgagcgca	cggggtcgge	ggcgatgtcg	gctacccacc	14880
cgacccgtct	tgaaacacgg	accaaggagt	ctaaccgtg	cgcgagtcag	gggctcgctcc	14940
gaaagccgcc	gtggcgcaat	gaagggtgaag	ggccccgccc	gggggcccga	gggtgggatcc	15000
cgaggcctct	ccagtccgcc	gagggcgccac	caccggcccg	tctcgcccg	cgcgccgggg	15060
aggtggagca	cgagcgtacg	cgttaggacc	cgaagatgg	tgaactatgc	ttgggcaggg	15120
cgaagccaga	ggaaactctg	gtggagggtcc	gtagcggtcc	tgacgtgcaa	atcggtcgctc	15180
cgacctgggt	atagggggcg	aagactaatc	gaaccatcta	gtagctgggt	ccctccgaag	15240
tttccctcag	gatagctggc	gctctcgctc	ccgacgtacg	cagttttatc	cggtaaagcg	15300
aatgattaga	ggctctgggg	ccgaaacgat	ctcaacctat	tctcaaactt	taaatgggta	15360
agaagcccg	ctcgctggcg	tggagccggg	cgtggaatgc	gagtgcctag	tgggcccactt	15420
ttggtaagca	gaactggcgc	tgcgggatga	accgaacgcc	gggttaaggc	gcccagatgcc	15480
gacgtcatc	agaccccaga	aaagggtgtg	gttgatatag	acagcaggac	ggtggccatg	15540
gaagtccgaa	tccgctaagg	agtgtgtaac	aactcacctg	ccgaatcaac	tagccctgaa	15600
aatggatggc	gctggagcgt	cgggcccata	cccgcccgctc	gccgcagtcg	gaacggaacg	15660
ggacggggagc	ggcccggggt	gcgcgtctct	cggggtcggg	ggtgctggg	ggggcccggt	15720
ccccgcctc	ccctccgcgc	gccgggttcg	cccccgcggc	gtcgggcccc	gcccagacctc	15780
cgccgcgacg	agtagggagg	ccgctcgcggt	gagccttgaa	gcctagggcg	cgggcccggg	15840
tggagccgcc	gcaggtggcg	atcttggtgg	tagtagcaaa	tattcaaacg	agaaacttga	15900
aggccgaagt	ggagaagggt	tccatgtgaa	cagcagttga	acatgggtca	gtcggtcctg	15960
agagatgggc	gagtgccgtt	ccgaagggac	gggagatggc	ctccgttgcc	ctcgcccgat	16020
cgaaggaggag	tcgggttcag	atccccgaat	ccggagtggc	ggagatgggc	gcccgcaggc	16080
cagtgcggta	acgcgaccga	tcccgagaa	gccggcgggg	ggcctcgggg	agagtctctct	16140
tttctttgtg	aaggccaggg	cgccctggaa	tgggttcgcc	ccgagagagg	ggcccggtgc	16200
ttggaaagcg	tcgcggttcc	ggcgccgtcc	ggtagctctc	cgctggccct	tgaaaatccg	16260
ggggagaggg	tgtaaatctc	gcgcggggcc	gtacccatat	ccgcagcagg	tctccaaggt	16320
gaacagcctc	tggcatgttg	gaacaatgta	ggtaagggaa	gtcggaagc	cggtaccgta	16380
aactcgggat	agggattggc	tctaagggtc	gggtcggtcg	ggctggggcg	ggaagcgggg	16440
ctggggcgcg	gccgcgggtg	gacgagggcg	cgccgccttc	tcccacgtcc	ggggagaccc	16500
cccgctcttt	ccgcccgggc	ccgcccctcc	ctctcccccg	cggggccccg	tcgtcccccg	16560
cgctcgccgc	ccccctcttc	ccccctcttc	cttcccgctc	gggggcccgg	cgggggtcgg	16620
cgcgccggcg	gggctccggg	gcggcggggtc	caaccccgcg	ggggttcggg	agcgggagga	16680
accagcggtc	cccggtgggg	cgggggggccc	ggacactcgg	ggggccggcg	gcggcgggcg	16740
ctctggacgc	cgcccgggcc	cttcccgtgg	atcgctcag	ctgcggcggg	cgtcgccgcc	16800
gctcccgggg	agcccggggg	gtgcccggcg	gggtcccttc	cccgcggggg	ctcgctccac	16860
ccccccatcg	cctctcccga	ggtgcgtggc	ggggcgggcg	gggcgtgtcc	cgcgcggtgtg	16920
gggggaacct	ccgcgtcggt	gttccccgcg	cgggtccgcc	ccccgggccc	cggttttcgc	16980
cgcggcgccc	ccgcctcggc	cgggcgctag	cagccgactt	agaactgggtg	cggaccaggg	17040
gaatccgact	gtttaattaa	aacaaagcat	cgcgaaaggc	cgcgccgggt	gttgacgca	17100
tgtgatttct	gcccagtgct	ctgaatgtca	aagtgaagaa	attcaatgaa	gcgcgggttaa	17160
acggcgggag	taactatgac	tctcttaagg	tagccaaatg	cctcgctcatc	taattatgta	17220
cgcgcatgaa	tggatgaacg	agattcccac	tgtccctacc	tactatccag	cgaaccacaa	17280
gccaaaggaa	cgggcttggc	ggaatcagcg	gggaaagaag	accctgttga	gcttgactct	17340
agtctggcac	ggtgaagaga	catgagaggt	gtagaataag	tgggaggccc	ccggcgcccg	17400
gccccgtcct	cgcgtcgggg	tcggggcacg	ccggcctcgc	gggcccggcg	tgaatatcca	17460
ctactctcat	cgttttttca	ctgacccgggt	gaggcggggg	ggcgagcccc	gaggggctct	17520
cgcttctggc	gccaagcgtc	cgctcccgcg	gtgcggggcg	gcgcgacccg	ctccggggac	17580
agtgccagggt	ggggagtttg	actggggcgg	tacacctgtc	aaacggtaac	gcaggtgtcc	17640
taaggcgagc	tcaggggagg	cagaaacctc	ccgtggagca	gaagggcaaa	agctcgcttg	17700
atcttgattt	tcagtagcaa	tacagaccgt	gaaagcgggg	cctcacgatac	cttctgacct	17760
tttgggtttt	aagcaggagg	tgtcagaaaa	gttaccacag	ggataactgg	cttgtggcgg	17820
ccaagcgttc	atagcgacgt	cgctttttga	tccttcgatg	tcggctcttc	ctatcattgt	17880
gaagcagaat	tcaccaagcg	ttggattgtt	cacccactaa	tagggaaacgt	gagctgggtt	17940
tagaccgtcg	tgagacaggt	tagttttacc	ctactgatga	tgtgttgttg	ccatggtaat	18000
cctgctcagt	acgagaggaa	ccgcaggttc	agacatttgg	tgtatgtgct	tggctgagga	18060
gccaatgggg	cgaagctacc	atctgtggga	ttatgactga	acgcctctaa	gtcagaatcc	18120
gccaagcgg	aacgatacgg	cagcgccgaa	ggagcctcgg	ttggcccccg	atagccgggt	18180
ccccgtccgt	cccgctcggc	ggggcccccg	cgtcgccccg	cgggcgccgc	gggtctcccc	18240
ccgcggggcg	tccggaccgg	ggtccggtgc	ggagagccgt	tcgtcttggg	aaacgggggtg	18300
cgccgggaaa	gggggcccgc	ctctcgcccc	tcacgttgaa	cgcacgttcg	tgtggaacct	18360
ggcgctaaac	cattcgtaga	cgacctgctt	ctgggtcggg	gtttcgtacg	tagcagagca	18420
gctccctcgc	tcgatctat	tgaagtcag	ccctcgacac	aagggtttgt	ctctcggggc	18480
tttcccgctc	cacgcccgtc	cgctcgccac	cgacctgtc	gccgcccggg	cgtcacgggg	18540
gcggctcgct	cgcccccgcc	gcgggtgccc	gaacgaccgt	gtgggtgggtg	ggggggggat	18600
cgtcttctcc	tccgtctccc	gaggacgggt	cgtttctctt	tecccttcgc	tcgctctcct	18660

-9-

tgggtgtggg	agcctcgtgc	cgtcgcgacc	gcggcctgcc	gtcgcctgcc	gccgcagccc	18720
cttgcctccc	ggccttggcc	aagccggagg	gcggaggagg	gggatcggcg	gcggcggcga	18780
ccgcggcgcg	gtgacgcacg	gtgggatccc	catcctcggc	gcgtccgtcg	gggacggccg	18840
gttggagggg	cgggaggggg	ttttcccgtg	aacgcgcggt	tcggcgccag	gcctctggcg	18900
gcgggggggg	cgctctctcc	gcccagagcat	ccccactccc	gccccctctc	ttcgcgcgcc	18960
gcggcgccga	cgtgcgtacg	aggggaggat	gtcgcgggtgt	ggaggcggag	aggggtccggc	19020
gcggcgccctc	ttccattttt	ttccccccaa	cttcggagggt	cgaccagtac	ttccggcgag	19080
actttgtttt	ttttttttcc	cccgatgctg	gagggtcgacc	agatgtccga	aagtgtcccc	19140
cccccccccc	ccccccggcg	cggagcggcg	gggacactct	ggactctttt	tttttttttt	19200
tttttttttt	ttaaattcct	ggaaacctta	ggtcgaccag	ttgtccgtct	tttactcctt	19260
catatagggtc	gaccagttact	ccgggtggta	ctttgtcttt	ttctgaaaat	cccagagggtc	19320
gaccagatat	ccgaaagtcc	ttctttttccc	tttactcttc	cccacagcga	ttctcttttt	19380
tttttttttt	tttgggtgtgc	ctcttttttga	cttatataca	tgtaaatagt	gtgtacgttt	19440
atatacttat	aggaggagggt	cgaccagttac	ttccggcgag	actttgtttt	tttttttttt	19500
ttccacgatg	atggagggtcg	accagatgtc	cgaaagtgtc	ccgtcccccc	ctcccccccc	19560
ccgcgacgcg	gcgggctcac	ttctggactct	tttttttttt	tttttttttt	tttaaatctc	19620
tggaacctta	aggtcgacca	gttgtccgtc	tttctactcat	tcatatagggt	cgaccgggtgg	19680
tactttgtct	ttttctgaaa	atcgcagagg	tcggaccagt	gtcagaaagt	ctgggtggctg	19740
ataaattatc	tgatctagat	ttgtttttct	gtttttcagt	tttgtgttgt	tttgtgttgt	19800
tttgtgttgt	tttgttttgt	tttgttttgt	tttgttttgt	tttgttttgt	tttgttttgt	19860
tttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	19920
gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	19980
tttgtgttgt	tttgtgttgt	tttgtgttgt	tttgtgttgt	tttgtgttgt	tttgtgttgt	20040
tacacaaaca	tgcaactttt	ttaaaataaa	tttttaaaat	aaatgcgaaa	atcgaccaat	20100
tatccctttc	cttctctctc	ttttttaaaa	attttctttg	tgtgtgtgtg	tgtgtgtgtg	20160
tgtgtgtgtg	tgctgtgtgt	tggtgtgtgt	cgtgcagcgt	gcgcgcgcgc	gttttataaa	20220
tacttataat	aatagggtcgc	cggtgtgtgt	tagcttcccg	gactccagag	gcagaggcag	20280
gcagacttct	gagttcgagg	ccagcctggg	ctacagagga	accctgtctc	gaaaaatgaa	20340
aataacatac	tacatacata	catacataca	tacatacata	catacataca	tacatatgag	20400
gttgacaggt	ttgcaatcct	ttagaatttt	gtttttaatt	aatgtgatag	agagatagat	20460
aatagataga	tggtatagagt	gatacaataa	taggtttttt	tttcagttaa	tatgagggtg	20520
attaaccact	tttccctttt	taggtttttt	tttttttccc	ctgtccatgt	gggtgtgtgg	20580
atttgaactc	aggaccctgg	cagggtcaact	ggaaaacgtg	ttttctatat	atataaatag	20640
tggtctgtct	gctgttttgt	tggtttgtgt	cttgtgtgt	tgcttgtctg	cttgtgtgt	20700
tgcttttttt	ttctctctga	gacagtattt	ctctgtgtaa	cctgggtgcc	tgaaactcac	20760
ttctgtagacc	agcctggcct	caatcgaaact	cagaaactct	cctgcctctt	gtctacctcc	20820
caatttttga	gtaaagggtg	gctacaccac	tgccctggcat	tattatcatt	atcattatta	20880
atatttattat	tagacagaac	gaaatcaact	agttgggtcct	gtttcgttaa	ttcatttgaa	20940
attagtgtga	ccaattagtt	ggctgggttg	ggagggttct	tttgtttccg	atttgggtgt	21000
ttgtggggct	ggggatcagg	tatctcaacg	gaatgcatga	aggttaagggt	gagatggctc	21060
gattttttgt	aagattactt	ttcttagtct	gaggaaaaaa	taaaataata	ttgggctacg	21120
tttctattgt	tcattttctt	ttctctttct	ttctttcttt	ctttcagata	aggagggtcg	21180
ccagttcctc	ctgccttctg	gaagatgtag	ctattgcatt	gggaaaagca	ttgtttgaga	21240
gatgtgctag	tgaaccagag	agtttggatg	tcaagccgta	taatgtttat	tacaatatag	21300
aaaagttcta	acaaaagtgt	ctttaacttt	tttttttttt	tttctccttc	tacttctact	21360
tggttctcact	ctgccaccaa	cgcgctttgt	acattgaatg	tgagctttgt	tttgcttaac	21420
agacatatat	tttttctttt	ggttttgctt	gacatgggtt	ccctttctat	ccgtgcaggg	21480
ttcccagacg	gcctttttgag	aataaaatgg	gaggccagaa	ccaaagtctt	ttgaataaag	21540
caccacaact	ctaactgttt	tggtctgttt	ccttcccaag	gcacagatct	ttccagcat	21600
ggaaaagcat	gtagcagttg	taggacacac	tagacgagag	caccagatct	cattgtgggt	21660
gggtgtgaac	caccaccat	gtgggtgcct	gggatttgaa	ctcaggatct	tcagaagacg	21720
agtcagggtc	ctaaaccgat	gagccatctc	ttccagccctc	ctacattcct	ttctaaggca	21780
tgaatgatcc	cagcatggga	agacagtctg	ccctctttgt	ggatatcac	catatactca	21840
ataaaaataat	gaaatgaatg	aagtctccac	gtatttattt	cttcgagcta	tctaaattct	21900
ctcacagcac	ctccccctcc	cccacactgc	ctttctccct	atgtttgggt	ggggctgggg	21960
gaggggtggg	gtgggggag	ggatctgcac	gtcttcttgc	aggtctgtga	actatttgcg	22020
atggcctggg	tctctgaact	gttgagcctt	gtctatccag	aggtgactg	gctagttttc	22080
tacctgaagt	ccctgagtg	tgatttccct	gtgaattc			22118

<210> 19  
 <211> 175  
 <212> DNA  
 <213> Mus musculus

<400> 19	ctccccgcgcg	gcccccggtgt	tcgcgcgttcc	cgtggcgcgcg	acaatgcgggt	tgtgcgtccca	60
	cgtgtgcgctg	tccgtgcaggt	cccggtgtgg	agtgccctcgc	tctcctcctc	ctccccggca	120



-10-

gcgttcccac gggtggggac caccggtgac ctgcacctct tcgggcctgg atccg 175

<210> 20  
 <211> 755  
 <212> DNA  
 <213> Mus musculus

<400> 20  
 ggtctggttg gaattgttga cctcgctctc ggggtgcggcc tttgggggaac ggcgggggtcg 60  
 gtcgtgcccc gcgcgcggac tgtgtcgggg ccacttccc gctcgagggg ggcgggtggcg 120  
 gcggcggttg tagtctcccc tgttgctctc tcccgggctc ttgggggggg tgccgtcggt 180  
 ttcggggccg gcgttgcttg gcttacgcag gcttggtttg ggactgcctc aggagtcgtg 240  
 ggcgggtgtga ttcccgccgg ttttgccctc cgtctgcctg ctttgccctc ggtttgcttg 300  
 gttcgtgtct cgggagcggt ggtttttttt ttttcgggt cccggggaga ggggtttttc 360  
 cgggggaagt tcccgtcgcc cctgcccgcc ggtgggtttt cgtttcgggc tgtgttcggt 420  
 tccccctccc cgtttcgccg tcggttctcc ccggtcggtc ggccctctcc ccggtcggtc 480  
 gcccggccgt gctgcgggac ccccccttct gggggggatg cccgggcacg cacgcgtccg 540  
 ggcggccact gtggtccggg agctgctcgg caggcgggtg agccagttgg aggggcgtca 600  
 tgcccccgcg ggctcccggt gccgacgcgg cgtgttcttt gggggggcct gtgcgtgcgg 660  
 gaaggctgcg cacgttgctg gtccttgcca gggaaagagg cttttttttt ttaggggggtc 720  
 gtcccttcgtc gtcccgctcg cgggtggatcc ggcct 755

<210> 21  
 <211> 463  
 <212> DNA  
 <213> Mus musculus

<400> 21  
 ggccgaggtg cgtctgcggg ttggggctcg tccggccccc tcgtcctccg ggaaggcggtt 60  
 tagcgggtac cgtcgccgcg ccgaggtggg cgcacgtcgg tgagataaac ccgagcggtg 120  
 ttctggttgt tggcgccggg ggctccggtc gatgtcttcc cctccccctc tccccgaggc 180  
 caggctcagcc tccgcctgtg ggcttcgtcg gccgtctccc cccccctcac gtccctcgcg 240  
 agcgagcccg tccgttcgac ctctcttccc ccttccccc atctttccgc gtcctcggtg 300  
 ccccggggtt ttcaacggcg ccccaacgtt cctccgctc tcgcgccgtg gtttggaacg 360  
 ctgggttccg tctccccgcc aaaccccggt tgggttggtc tccggccccc gcttgctctt 420  
 cgggtctccc aacccccggc cgggaagggtt cgggggttcc ggg 463

<210> 22  
 <211> 378  
 <212> DNA  
 <213> Mus musculus

<400> 22  
 ggattcttca ggattgaaac ccaaaccggt tcagtttctt ttccggctcc ggccggggggg 60  
 ggccggcccc ggccggtttg tgagttagat aacctcgggc cgatcgacg ccccccggtg 120  
 cggcgacgac ccattcgaac gtctgcctta tcaactttcg atggtagtcg atgtgcctac 180  
 catggtgacc acgggtgacg gggaatcagg gttcgattcc ggagagggag cctgagaaac 240  
 ggctaccaca tccaaggaag gcagcaggcg cgcaaattac ccactccga cccggggagg 300  
 tagtgacgaa aaataacaat acaggactct ttcgaggccc tgtaattgga atgagtcac 360  
 tttaaatcct ttaagcag 378

<210> 23  
 <211> 378  
 <212> DNA  
 <213> Mus musculus

<400> 23  
 gatccattgg agggcaagtc tggtgccagc agccgcggta attccagctc caatagcgta 60  
 tattaagatt gctgcagtta aaaagctcgt agttggatct tgggagcggg cgggcgggtc 120  
 gccgcgagge gagtcaaccg ccgtccccgc cccttgccctc tcggcgcccc ctcgatgtc 180  
 ttagctgagt tgtcccgcg ggcccgaaag gtttactttg aaaaaattag agttgtttca 240  
 aagcaggccc gagccgcctg gataccgcca gctaggaaat aatggaatag gaccgcgggt 300  
 cctattttgt ttggttttct gaactgagcc catgattaag ggaaacggcc gggggcattc 360  
 ccttattgcg cccccccta 378

<210> 24  
 <211> 719



-11-

&lt;212&gt; DNA

&lt;213&gt; Mus musculus

&lt;400&gt; 24

ggatctttcc	cgctccccgt	tcctcccggc	ccctccaccc	gcgctctcc	ccctttcttt	60
tccctctcc	ggaggggggg	gaggtggggg	cgctggggcg	gggtcggggg	tggggtcggc	120
gggggaccgc	ccccggccgg	caaaaggccg	cgccggggcg	cacttcaacc	gtagcgggtc	180
gccgcgaccg	gctacgagac	ggctgggaag	gcccgcgggg	gaatgtggct	cggggggggc	240
ggcgcgtctc	agggcgcgcc	gaaccacctc	accccgagtg	ttacagccct	ccggcccgcg	300
tttcgcgga	tcccggggcc	gaggggaagc	cgatacccg	tcgcccgcgt	tttccctcc	360
ccccgtccgc	ctcccggggc	ggcgtggggg	tggggggccg	gccgccctc	ccacgcccgt	420
ggtttctctc	tctcccggtc	tcggccgggt	tggggggggg	agcccggttg	ggggcggggg	480
ggactgtcct	cagtgcgccc	cgggcgtcgt	cgcccgctcg	ggcccggggg	gttctctcgg	540
tcacgccgcg	cccgacgaag	ccgagcgcac	ggggtcggcg	gcgatgtcgg	ctacccaccc	600
gaccgcgtct	gaaacacgga	ccaaggagtc	taacgcgtgc	gcgagtcagg	ggctcgcacg	660
aaagccgccc	tggcgcaatg	aaggtgaagg	gccccgtccg	ggggcccag	gtgggatcc	719

&lt;210&gt; 25

&lt;211&gt; 685

&lt;212&gt; DNA

&lt;213&gt; Mus musculus

&lt;400&gt; 25

cgaggcctct	ccagtcggcc	gagggcgcac	caccggcccc	tctcgccgcg	cgcgtegggg	60
aggtggagca	cgagcgtacg	cgtaggacc	cgaaagatgg	tgaactatgc	ctgggcaggg	120
cgaagccaga	ggaaactctg	gtggaggtcc	gtagcggctc	tgacgtgcaa	atcggtcgte	180
cgacctgggt	ataggggcga	aagactaatc	gaaccatcta	gtagctgggt	ccctccgaag	240
tttccctcag	gatagctggc	gctctcgcaa	ccttcggaag	cagttttatc	cgggtaaagg	300
cggaatggat	taggaggtct	tggggccgga	aacgatctca	aactatttct	caaactttaa	360
atgggtaagg	aagcccggct	cgctggcggt	gagccggggc	tggaaatgca	gtgcctagtg	420
ggccactttt	ggtaagcaga	actggcgctg	cgggatgaac	cgaacgccgg	gttaaggcgc	480
ccgatgccga	cgctcatcag	acccagaaaa	aggtgttggt	tgatatagac	agcaggacgg	540
tggccatgga	agtcgggaatc	cgctaaggag	tgtgtaacaa	ctcacctgcc	gaatcaacta	600
gccctgaaaa	tggatggcgc	tggagcgtcg	ggcccatacc	cggccgtcgc	cggcagtcgg	660
aacgggacgg	gacggggagcg	gccgc				685

&lt;210&gt; 26

&lt;211&gt; 5162

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Chimeric bacterial plasmid

&lt;400&gt; 26

gacggatcgg	gagatctccc	gatccccctat	ggctcgactct	cagtacaatc	tgctctgatg	60
ccgcatagtt	aagccagtat	ctgctccctg	cttgtgtgtt	ggaggtcgct	gagtagtgcg	120
cgagcaaaat	ttaagctaca	acaaggcaag	gcttgaccga	caattgcatg	aagaatctgc	180
ttagggttag	gcgttttgcg	ctgcttcgcg	atgtacgggc	cagatatacg	cgttgacatt	240
gattattgac	tagttattaa	tagtaatcaa	ttacggggtc	attagttcat	agcccatata	300
tggagttccg	cgttacataa	cttacggtaa	atggcccggc	tggctgaccg	cccaacgacc	360
cccgcccat	gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	420
attgacgtca	atgggtggac	tatttacgggt	aaactgcccc	cttggcagta	catcaagtgt	480
atcatatagg	aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	540
atgcccagta	catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	600
tcgctattac	catgggtgat	cggtttttggc	agtagatcaa	tgggcgtgga	tagcgggttg	660
actcacgggg	atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	720
aaaatcaacg	ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatggggc	780
gtaggcgtgt	acgggtgggag	gtctatataa	gcagagctct	ctggctaact	agagaaccca	840
ctgcttactg	gcttatcgaa	attaatacga	ctcactatag	ggagacccaa	gcttgggtacc	900
gagctcggat	cgatatctgc	ggccgcgtcg	acggaattca	gtggatccac	tagtaacggc	960
cgccagtggt	ctggaattaa	ttcgctgtct	cgaggggcca	gctgttgggg	tgagtactcc	1020
ctctcaaaag	cgggcatgac	ttctgcgcta	agattgtcag	tttccaaaaa	cgaggaggat	1080
ttgatattca	cctggcccgc	ggtagtcctt	ttgaggttgg	ccgcgtccat	ctggtcagaa	1140
aagacaatct	ttttgtttgc	aagcttgagg	tgtggcaggc	ttgagatctg	gccatacact	1200
tgagtgacaa	tgacatccac	tttgcccttc	tctccacagg	tgtccactcc	caggtccaac	1260
tgcaggtcga	gcatgcatct	agggcggcca	attccgcccc	tctccctccc	ccccccctaa	1320

-12-

cgttactggc	cgaagccgct	tggaaataagg	cgggtgtgcg	tttgtctata	tgtgattttc	1380
caccatattg	ccgtcttttg	gcaatgtgag	ggcccggaag	cctggccctg	tcttcttgac	1440
gagcattcct	aggggtcttt	cccctctcgc	caaaggaatg	caaggtctgt	tgaatgtcgt	1500
gaaggaagca	gttccctctg	aagcttcttg	aagacaaaca	acgtctgtag	cgaccctttg	1560
caggcagcgg	aacccccccac	ctggcgacag	gtgectctgc	ggccaaaagc	cacgtgtata	1620
agatacacct	gcaaaggcgg	cacaacccca	gtgccacgtt	gtgagttgga	tagttgtgga	1680
aagagtcaaa	tggctctcct	caagcgtatt	caacaagggg	ctgaaggatg	cccagaaggt	1740
accccatgtg	atgggatctg	atctggggcc	tgggtgcaca	tgctttacat	gtgtttagtc	1800
gaggttaaaa	aaacgtctag	gccccccgaa	ccacggggac	gtggttttcc	tttgaaaaac	1860
acgatgataa	gcttgccaca	acccgggatc	caccgggtcg	caccatgggtg	agcaaggcgg	1920
aggagctggt	caccgggggtg	gtgcccattc	tggctcgagct	ggacggcgac	gtaaacggcc	1980
acaagttcag	cgtgtccggc	gagggcgagg	gcgatgccac	ctacggcaag	ctgaccctga	2040
agttcatctg	caccaccggc	aagctgcccg	tgccctggcc	caccctcgtg	accaccctga	2100
cctacggcgt	gcagtgcttc	agccgttacc	ccgaccacat	gaagcagcac	gacttcttca	2160
agtcggccat	gcccgaagge	tacgtccagg	agcgaccat	cttcttcaag	gacgacggca	2220
actacaagac	ccgcgcggag	gtgaagttcg	agggcgacac	cctgggtgaac	cgcacogagc	2280
tgaaggacat	cgacttcaag	gaggacggga	acatcctggg	gcacaagctg	gagtacaact	2340
acaacagcca	caactcttat	atcatggcgg	acaagcagaa	gaacggcatc	aaggtgaact	2400
tcaagatccg	ccacaacatc	gaggacggga	gcgtgcagct	cgccgaccac	taccagcaga	2460
acacccccat	cggcgacggc	cccgtgctgc	tgcccgacaa	ccactacctg	agcaccagtc	2520
ccgcccctgag	caagaccccc	aacgagaagc	gcgatccat	ggtcctgctg	gagttcgtga	2580
ccgcgcgcgg	gatcactctc	ggcatggacg	agctgtacaa	gtaaagcggc	cctagagctc	2640
gctgatcagc	ctcgactgtg	cctctagtgt	ccagccatct	gttggttgcc	cctccccctg	2700
gccttctctg	accctggaag	gtgccactcc	cactgtcctt	tcctaataaa	atgaggaaat	2760
tgcatcgcat	tgtctgagta	ggtgtcattc	tattctgggg	ggtgggggtg	ggcaggacag	2820
caagggggag	gatttgggaag	acaatagcag	gcattgctgg	gatgcggtgg	gctctatggc	2880
ttctgaggcg	gaaagaacca	gttggggctc	gagtgcatte	tagttgtggg	ttgtccaaac	2940
tcattcatgt	atcttatcat	gtctgtatc	cgctgcacct	tagctagagc	ttggcgtaat	3000
catggtcata	gctgtttcct	gtgtgaaatt	gttatccgct	cacaattcca	cacaacatac	3060
gagccggaag	cataaagtgt	aaagcctggg	gtgcctaatg	agttagctaa	ctcacattaa	3120
ttgcgttgcg	ctcactgccc	gctttccagt	cgggaaacct	gtcgtgccag	ctgcattaat	3180
gaatcggcca	acgcgcgggg	agaggcgggt	tgcgtattgg	gcgctcttcc	gcttccctgc	3240
tcactgactc	gctgcgctcg	gtcgttcggc	tgccgcgagc	ggtatcagct	cactcaagg	3300
cggtaatacg	ggttatccaca	gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaag	3360
gccagcaaaa	ggccaggaac	cgtaaaaagg	ccgcgttgct	ggcgtttttc	cataggctcc	3420
gcccccttga	cgagcatcac	aaaaatcgac	gctcaagtc	gaggtggcga	aacccgacag	3480
gactataaag	ataccaggcg	tttccccctg	gaagctccct	cgtgcgctct	cctgtttccga	3540
ccctgccgct	taccggatac	ctgtccgcct	ttctcccttc	gggaagcgtg	gcgctttctc	3600
aatgctcacg	ctgtaggtat	ctcagttcgg	tgtaggtcgt	tcgctccaag	ctgggctgtg	3660
tgacgaacc	ccccgttcag	cccgaccgct	gcgccttact	cggttaactat	cgtcttgagt	3720
ccaacccggt	aagacacgac	ttatcgccac	tgccagcagc	cactggtaac	aggattagca	3780
gagcgaggta	tgtaggcggg	gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	3840
ctagaaggac	agtatttggg	atctgcgctc	tgctgaagcc	agttaccttc	ggaaaaagag	3900
ttggtagctc	ttgatccggc	aaacaaacca	ccgctggtag	cggtgggttt	tttgtttgca	3960
agcagcagat	tacgcgcaga	aaaaaaggat	ctcaagaaga	tcctttgatc	ttttctacgg	4020
ggtctgacgc	tcagtggaaac	gaaaactcac	gttaagggat	tttggctcatg	agattatcaa	4080
aaaggatctt	cacctagatc	cttttaaatt	aaaaatgaag	ttttaaatca	atctaaagta	4140
tatatgagta	aacttgggtc	gacagttacc	aatgcttaat	cagtgaggca	cctatctcag	4200
cgatctgtct	atttcgttca	tcctagtttg	cctgactccc	cgtcgtgtag	ataactacga	4260
tacgggaggg	cttaccatct	ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	4320
cggtccagga	tttatcagca	ataaaccagc	cagccgggaag	ggccgagcgc	agaagtgggt	4380
ctgcaacttt	atccgcctcc	atccagtcta	tttaattgttg	ccgggaagct	agagtaagta	4440
gttcgccaggt	taatagtttg	cgcaacgctg	ttgccattgc	tacaggcatc	gtgggtgtcac	4500
gctcgtcggt	tggtatggct	tcattcagct	ccggttccca	acgatcaagg	cgagttacat	4560
gatcccccat	gttgtgcaaa	aaagcggtta	gtcccttcgg	tcctccgac	gttgtcagaa	4620
gtaagttggc	cgcagtggtt	tcactcatgg	ttatggcagc	actgcataat	tctcttactg	4680
tcattgccatc	cgtaagatgc	ttttctgtga	ctgggtgagta	ctcaaccaag	tcattctgag	4740
aatagtgtat	cggcgacccg	agttgctctt	gcggcgctgc	aatacgggat	aatccgcgc	4800
cacatagcag	aactttaaaa	gtgctcatca	ttggaaaacg	ttcttcgggg	cgaaaactct	4860
caaggatctt	accgctgttg	agatccagtt	cgatgtaaac	cactcgtgca	cccaactgat	4920
cttcagcattc	ttttactttc	accagcgttt	ctgggtgagc	aaaaacagga	aggcaaatg	4980
ccgcaaaaaa	gggaataagg	gcgacacgga	aatgttgaat	actcatactc	ttcctttttc	5040
aatattattg	aagcattttat	caggggttatt	gtctcatgag	cggatacata	tttgaatgta	5100
tttagaaaaa	taaacaaata	gggggttcgc	gcacatttcc	ccgaaaagtg	ccacctgacg	5160
tc						5162

-13-

<211> 5627  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> pMG plasmid from InvivoGen; IRES sequence modified  
EMCV nucleotides 2736-3308

<400> 27  
caccggcgaa ggaggcctag atctatcgat tgtacagcta gctcgacatg ataagataca 60  
ttgatgagtt tggacaaaacc acaactagaa tgcagtgaaa aaaatgcttt atttgtgaaa 120  
tttgtgatgc tattgcttta tttgtgaaat ttgtgatgct attgctttat ttgtaaccat 180  
tataagctgc aataaacaag ttaacaacaa caattgcatt catttttatgt ttccagggttca 240  
gggggagggtg tgggagggttt tttaaagcaa gtaaaacctc tacaaatgtg gtagatccat 300  
ttaaattgta attaagaaca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa 360  
ggcgcggttg ctggcggtttt tccataggct cgcgcgcctc gacgagcatc acaaaaaatcg 420  
acgctcaagt cagagggtggc gaaacccgac aggaactata agataccagg cgtttccccc 480  
tggaaagctcc ctgcgtgcgt ctccgtgttc gaccctgcgc ctaccggat acctgtccgc 540  
ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggt atctcagttc 600  
gggtgaggte gttcgtcca agctgggctg tgtgcacgaa ccccccgttc agcccgaccg 660  
ctgcgcctta tccggtaact atcgtcttga tgtccaaccg gtaagacacg acttatcgcc 720  
actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 780  
gttcttgaag tgggtggccta actacggcta cactagaaga acagtatttg gtatctgccc 840  
tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaaacaaac 900  
caccgctggg agcgggtgggt tttttgtttg caagcagcag attacgcgca gaaaaaaagg 960  
atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtgga acgaaaaactc 1020  
acgttaagggt attttgttca tggctagtta attaagctgc aataaacaat cattattttc 1080  
attggatctg tgtgttgggt ttttgtgtgg gcttggggga gggggaggcc agaattgactc 1140  
caagagctac aggaaggcag gtcagagacc ccactggaca aacagtgggt ggactctgca 1200  
ccataacaca caatcaacag gggagtggag tggatcgagc tagagtccgt tacataactt 1260  
acggtaaatg gccgcgctgg ctgaccgccc aacgaccccc gccatttgac gtcaataatg 1320  
acgtatgttc ccatagtaac gccaatagggt actttccatt gacgtcaatg ggtggagtat 1380  
ttacggtaaa ctgccactt ggcagtaact caagtgtatc atatgccaag taacccccct 1440  
attgacgtca atgacggtaa atggcccgc tggcattatg ccagtagcat gaccttatgg 1500  
gactttccta ctgggcagta catctacgta ttagtcatcg ctattaccat ggtgatgcgg 1560  
ttttggcagt ccatcaatgg gcgtggatag cggtttgact cacggggatt tccaagtctc 1620  
caccgccattg acgtcaatgg gagtttgttt tggcaccaaa atcaacggga ctttccaaaa 1680  
tgtcgtaaca actccgcccc attgacgcaa atggggcggt ggcgtgtacg gtgggagggtc 1740  
tatataagca gacgtcggtt agtgaaccgt cagatcgccg ggagacgcca tccagctgt 1800  
tttgacctcc atagaagaca ccgggaccga tccagcctcc gcggccggga acggtgcatt 1860  
ggaaacggga tccccgtgc caagagtgc gtaagtaccg cctatagagt ctataggccc 1920  
accccttgg atgttatgct atgtataact gtttttggct tggggtctat acacccccgc 1980  
ttcctcatgt tataggatgat ggtatagctt agcctatagg tgtgggttat tgaccattat 2040  
tgaccactcc cctattgggt acgatacttt ccattactaa tccataacat ggctctttgc 2100  
cacaactctc tttattggct atatgccaat acactgtctc tcagagactg acacggactc 2160  
tgtattttta caggatgggg tctcatttat tatttcaaaa ttcacatata caacaccacc 2220  
gtccccagtg cccgcagttt ttattaaaca taacgtggga tctccacgcg aatctcgggt 2280  
acgtgttccg gacatgggct cttctccggt agcggcgagg cttctacatc cgagccctgc 2340  
tcccatgcct ccagcgactc atggctcgctc ggcagctcct tgctcctaac agtggaggcc 2400  
agacttaggc acagcacgat gccaccacc accagtgtgc cgcacaagge cgtggcggtg 2460  
gggtatgtgt ctgaaaatga gctcggggag cgggcttgca ccgctgacgc atttgggaaga 2520  
cttaaggcag cggcagaaga agatgcaggc agctgagttg ttgtgttctg ataagagtca 2580  
gaggttaact ccgttgcggt gctgttaacg gtggagggca gtgtagtctg agcagttact 2640  
gttgctgcg cgcgcgccac cagacataat agctgacaga ctaacagact gttcctttcc 2700  
atgggtcttt ctgacagtca cccgggggat ccttcgaacg tagctctaga ttgagtcgac 2760  
gttactggcc gaagccgctt ggaataagge cgggtgtcgt ttgtctatat gttattttcc 2820  
accatattgc cgtcttttgg caatgtgagg gcccggaac aagggtctgt gaatgtcgtg 2880  
agcattccta ggggtctttc cctctcgcc aaaggaatgc aaggtctgtt gaatgtcgtg 2940  
aaggaaagcag ttcctctgga agcttcttga agacaaacaa cgtctgtagc gaccttttgc 3000  
aggcagcgga acccccacc tggcgacagg tgccctctgc tgagttggat agttgtggaa 3060  
gatacacctg caaaggcggc acaacccag tgccacgttg tgaaggtatg ccagaaggta 3120  
agagtcfaat ggctctcctc aagcgtatct aacaaggggc tgaaggtatg ccagaaggta 3180  
ccccattgta tgggacttga tctggggcct cgtgtcacat gctttacatg tgttttagtcg 3240  
aggttaaaaa aacgctcagg ccccccgaac cagggggacg tgggttttcc ttgaaaaaca 3300  
cgataatacc atgggtaagt gatattctact agttgtgacc ggcgcctagt gttgacaatt 3360  
aatcatcggc atagtatatc ggcatagtat aatacagact actataggag ggcaccatg 3420  
tcgactacta acctcttctc gtttcttaca gctgagatca ccggtaggag ggccatcatg 3480

-14-

aaaaagcctg	aactcaccgc	gacgtctgtc	gcgaagtttc	tgatcgaaaa	gttcgacagc	3540
gtctccgacc	tgatgcagct	ctcggagggc	gaagaatctc	gtgctttcag	cttcgatgta	3600
ggagggcgctg	gatatgtcct	gcgggtaaat	agctgcgcgc	atgggtttcta	caaagatcgt	3660
tatgtttatc	ggcactttgc	atcggccgcg	ctcccgatcc	cgggaagtgt	tgacattggg	3720
gaattcagcg	agagcctgac	ctattgcatc	tcccgcctgt	cacagggtgt	cacgttgcaa	3780
gacctgcctg	aaaccgaact	gcccgcctgt	ctgcaacccg	tcgcggagct	catggatgctg	3840
atcgctgcgg	ccgatcttag	ccagacgagc	gggttcggcc	cattcggacc	gcaaggaatc	3900
gggtcaatata	ctacatggcg	tgatttcata	tgccgcgattg	ctgatcccca	tgtgtatcac	3960
tgggcaactg	tgatggacga	caccgtcagt	gcgtccgtcg	cgcaggctct	cgatgagctg	4020
atgctttggg	ccgaggactg	ccccgaagtc	cggcacctcg	tgacgcggga	tttcggctcc	4080
aacaatgtcc	tgacggacaa	tgccgcgata	acagcgttca	ttgactggag	cgaggcgatg	4140
ttcgggggatt	cccaatacga	ggtcgccaac	atcttcttct	ggaggccgtg	gttggcttgt	4200
atggagcagc	agacgcgcta	cttcgagcgg	aggcatccgg	agcttgccag	atcgcccgccg	4260
ctccggggcgt	atatgtcccg	cattggctct	gaccaactct	atcagagctt	ggttgacggc	4320
aatttcgatg	atgcagcttg	ggcgagggtt	cgatgcgacg	caatcgcccg	atccggagcc	4380
gggactgtcg	ggcgtagaca	aatcgcccg	agaagcgccg	ccgtctggac	cgatggctgt	4440
gtagaagtac	tcgcccagatg	tggaaccgga	cgccccagca	ctcgtccgag	ggcaaaggaa	4500
tgagtgcgaga	atctgctaga	gggccttatt	ctatagtgtc	acctaataatg	tagagctcgc	4560
tgatcagcct	cgactgtgcc	ttctagtgtc	cagccatctg	ttgtttgccc	ctcccccggtg	4620
ccttccttga	ccctggaagg	tgccactccc	actgtccttt	cctaataaaa	tgaggaaatt	4680
gcacgcgatt	gtctgagtag	gtgtcattct	attctggggg	gtgggtggg	gcaggacagc	4740
aagggggagg	attgggaaga	caatagcagg	catgcgcagg	gcccaattgc	tcgagcggcc	4800
gcaataaaat	atctttatct	tcattacatc	tgtgtgttgg	ttttttgtgt	gaatcgtaac	4860
taacatacgc	tctccatcaa	aacaaaacga	tctatcgaag	actagcaaaa	taggctgtcc	4920
ccagtgcgaag	tgacaggtgc	agaacatttc	cagtcccccg	gatctgcgat	cgctccgggtg	4980
cccgtcagtg	ggcagagcgc	acatcgcccc	gcggggtaaa	gaagtggggg	ggagggttcg	5040
gcaattgaac	cggtgcctag	agaaggtggc	gagaaaccgta	ctgggaaagt	gatgtcgtgt	5100
actggctccg	cttttttccc	gaggggtggg	ccagaacaca	tataagtgc	gtagtcccg	5160
tgaacgttct	ttttcgcaac	gggttttgccg	cagcagccca	gctgaagctt	cgaggggctc	5220
gcactctctc	ttcacgcgcg	cgccgcctta	cctgaggccg	ccatccacgc	cggttgagtc	5280
gcgttctgcc	gcctcccgcg	tggtgtgctt	cctgaactgc	gtccgcctgc	taggttaagt	5340
taaagctcag	gtcgagaccg	ggcctttgtc	cggcgctccc	ttggagccta	cctagactca	5400
gccggctctc	cacgctttgc	ctgaccctgc	ttgtcactac	ctacgtcttt	gtttcgtttt	5460
ctgttctgcg	ccgttacaga	tccaagctgt	gaccggcgcc	tacgtaagt	atatctacta	5520
gatttatcaa	aaagagtgtt	gacttgtgag	cgctcacaat	tgatacttag	attcatcgag	5580
agggacacgt	cgactactaa	ccttcttctc	tttctacag	ctgagat		5627

&lt;210&gt; 28

&lt;211&gt; 553

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pMG plasmid from InvivoGen: EMCV IRES sequence

&lt;400&gt; 28

aacgttactg	gccgaagccg	cttgggaataa	ggccgggtgtg	cgtttgtcta	tatgttattt	60
tccaccatat	tgccgtcttt	tggcaatgtg	agggcccggga	aacctggccc	tgtcttcttg	120
acgagcattc	ctaggggtct	ttcccctctc	gccaaaggaa	tgcaaggctc	gttgaatgtc	180
gtgaaggaag	cagttcctct	ggaagcttct	tgaagacaaa	caacgtctgt	agcgaccctt	240
tgcaggcagc	ggaacccccc	acctggcgac	aggtgcctct	gcggccaaaa	gccacgtgta	300
taagatacac	ctgcaaaggc	ggcacaaccc	cagtgcacacg	ttgtgagttg	gatagttgtg	360
gaaagagtca	aatggctctc	ctcaagcgta	ttcaacaagg	ggctgaagga	tgcccagaag	420
gtaccccat	gtatgggac	tgatctgggg	cctcggtgca	catgctttac	tgtgtgttag	480
tcgaggttaa	aaaacgtcta	ggccccccga	accacgggga	cgtgggtttc	ctttgaaaaa	540
cacgatgata	ata					553

&lt;210&gt; 29

&lt;211&gt; 4692

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pDSred1-N1 plasmid from Clontech

&lt;400&gt; 29

tagttattaa tagtaatcaa ttacgggggtc attagttcat agcccatata tggagttccg 60

-15-

cgttacataa	cttacggtaa	atggccccgc	tggctgaccg	cccaacgacc	cccgccatt	120
gacgtcaata	atgacgtatg	ttcccatagt	aacgcccaata	gggactttcc	attgacgtca	180
atgggtggag	tattttacgg	aaactgcccc	cttggcagta	catcaagtgt	atcatatgcc	240
aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgcccagta	300
catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac	360
catggtgatg	cggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	420
atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	480
ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	gtaggcgtgt	540
acgggtgggag	gtctatataa	gcagagctgg	tttagtgaa	cgtcagatcc	gctagcgcta	600
ccggactcag	atctcgagct	caagcttcga	attctgcagt	cgacgggtacc	gcgggcccg	660
gatccaccgg	tcgccaccat	ggtgcgctcc	tccaagaacg	tcacaaagga	gttcatgcgc	720
ttcaagggtgc	gcatggaggg	caccgtgaac	ggccacgagt	tcgagatcga	gggcgagggc	780
gagggccgccc	cctacgaggg	ccacaacacc	gtgaagctga	aggtgaccaa	gggcccggccc	840
ctgcccctcg	cctgggacat	cctgtccccc	cagttccagt	acggctccaa	ggtgtacgtg	900
aagcaccgcc	ccgacatccc	cgactacaag	aagctgctcc	tccccgaggg	cttcaagtgg	960
gagcgcgtga	tgaacttcga	ggacggcggc	gtgggtgaccg	tgaccagga	ctcctccctg	1020
caggacggct	gcttcatcta	caaggtgaag	ttcatcgccg	tgaacttccc	ctccgacggc	1080
cccgtaatg	agaagaagac	catgggctgg	gagggctcca	ccgagcgcc	gtaccccgc	1140
gacggcgctgc	tgaagggcga	gatccacaag	gccctgaagc	tgaaggacgg	cggccactac	1200
ctggtggagt	tcaagtccat	ctacatggcc	aagaagcccg	tgacgctgcc	cggctactac	1260
tacgtggact	ccaagctgga	catcacctcc	cacaacgagg	actacacccat	cgtgagcctt	1320
tacgagcgca	ccgagggccg	ccaccacctg	ttcctgtagc	ggccgcgact	ctagatcata	1380
atcagccata	ccacatttgt	agaggtttta	cttgccttaa	aaaacctccc	acacctcccc	1440
ctgaacctga	aacataaaaat	gaatgcaatt	gttgtgtgta	acttgtttat	tgacgcttat	1500
aatgggttaca	aataaagcaa	tagcatcaca	aatttccaaa	ataaagcatt	tttttctactg	1560
cattctagtt	gtggtttgtc	caaactcatc	aatgtatctt	aaggcgtaaa	ttgtaaagct	1620
taatattttg	ttaaaattcg	cgttaaaatt	ttgttaaatt	agctcatttt	ttaaaccaata	1680
ggccgaaatc	ggcaaaatcc	cttataaatc	aaaagaatag	accgagatag	ggttgagtgt	1740
tgttccagtt	tggaaacaaga	gtccactatt	aaagaacgtg	gactccaacg	tcaaagggcg	1800
aaaaaccgtc	tacagggcg	atggcccact	acgtgaacca	tcaccctaata	caagtttttt	1860
ggggtcgagg	tggcgtaaag	cactaaatcg	gaaccctaaa	gggagccccc	gatttagagc	1920
ttgacgggga	aagccggcga	acgtggcgag	aaaggaaggg	aagaaaacga	aaggagccgg	1980
cttgaggggc	ctggcaagtg	tagcgggtcac	gctgcgtcga	accaccacac	ccgccgct	2040
taatgcggcg	ctacagggcg	cgctcaggtg	cacttttccg	ggaaatgtgc	gcggaacccc	2100
tatttgttta	tttttctaaa	tacattcaaa	tatgtatcga	ctcatgagac	aataaccctg	2160
ataaatgatt	caataatatt	gaaaaaggaa	gagtcctgag	gcggaagaa	ccagctgtgg	2220
aatgtgtgtc	agttagggtg	tggaaagtcc	ccaggctccc	cagcaggcag	aagtatgcaa	2280
agcatgcatc	tcaattagtc	agcaaccagg	tgtggaaagt	ccccaggctc	cccagcaggc	2340
agaagtatg	aaagcatgca	tctcaattag	tcagcaacca	tagtcccgc	cctaactccg	2400
cccatccgc	ccctaactcc	gcccagttcc	gcccattctc	cgccccatgg	ctgactaatt	2460
ttttttat	atgcagaggc	cgaggccgcc	tcggcctctg	agctattcca	gaagtagtga	2520
ggaggcttt	ttggaggcct	aggcttttgc	aaagatcgat	caagagacag	gatgaggatc	2580
gtttgcgatg	attgaacaag	atggatttga	cgcagggtct	ccggccgctt	gggtggagag	2640
gctattcgcc	tatgactggg	cacaacagac	aatcggctgc	tctgatgccg	ccgtgttccg	2700
gctgtcagcg	caggggccc	cggttctttt	tgtaagacc	gacctgtccg	gtgccctgaa	2760
tgaactgcaa	gacgaggcag	cgcggctatc	gtggctggcc	acgacgggcg	ttccttgcgc	2820
agctgtgctc	gacgttgtca	ctgaagcggg	aagggactgg	ctgctattgg	gcgaagtgcc	2880
ggggcaggat	ctcctgtcat	ctcaccttgc	tcctgccgag	aaagtatcca	tcattggctga	2940
tgcaatgcgg	cggctgcata	cgcttgatcc	ggctacctgc	ccattcgacc	accaagcgaa	3000
acatcgcatc	gagcagacac	gtactcggat	ggaagccggg	cttgtcgatc	aggatgatct	3060
ggacgaagag	catcaggggc	tcgcgccagc	cgaactgttc	gccaggctca	aggcgagcat	3120
gcccagcggc	gaggatctcg	tcgtgaccca	tggcgaatgc	tgcttgccga	atatcatggg	3180
ggaaaatggc	cgcttttctg	gattcatcga	ctgtggccgg	ctgggtgtgg	cggaccgcta	3240
tcaggacata	gcgttggcta	cccgtgata	tgctgaagag	cttggcggcg	aatgggctga	3300
ccgcttcctc	gtgctttacg	gtatcgccgc	tcocgattcg	cagcgcacgc	ccttctatcg	3360
ccttcttgac	gagttcttct	gagcgggact	ctggggttcg	aatgaccga	ccaagcgacg	3420
cccaacctgc	catcacgaga	tttccgattcc	accgcccgct	tctatgaaag	gttgggcttc	3480
ggaatcgctt	ctcgggacgc	cggctggatg	atcctccagc	gcgggggatct	catgtctggag	3540
ttcttcgccc	accctagggg	gaggctaact	gaaacacgga	aggagacaat	accggaagga	3600
acccgcgcta	tgcaggcaat	aaaaagacag	aataaaacgc	acggtgttgg	gtcgtttgtt	3660
cataaacgcg	gggttcggtc	ccagggctgg	cactctgtcg	ataccccacc	gagaccccat	3720
tggggccaat	acgcccgcgt	ttcttccttt	tccccacccc	accccccaag	ttcgggtgaa	3780
ggccaggggc	tcgcagccaa	cgtcggggcg	gcaggccctg	ccatagcctc	agggtactca	3840
tatatacttt	agattgattt	aaaacttcat	ttttaattta	aaaggatcta	ggtgaagatc	3900
ctttttgata	atctcatgac	caaaatccct	taacgtgagt	tttctgtcca	ctgagcgtca	3960
gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	ttttctcg	cgtaatctgc	4020
tgcttgcaaa	caaaaaaacc	accgctacca	gcgggtggtt	gtttgcccga	tcaagagcta	4080

-16-

ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	tactgtcctt	4140
ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcaccgcc	tacatacctc	4200
gctctgctaa	tcctgttacc	agtggctgct	gccagtggcg	ataagtctgt	tcttaccggg	4260
ttggactcaa	gacgatagtt	accggataag	gcgagcggt	cgggctgaac	gggggggtcg	4320
tgcacacagc	ccagcttgga	gcgaacgacc	tacaccgaac	tgagatacct	acagcgtgag	4380
ctatgagaaa	gcgccacgct	tcccgaaagg	agaaaggcgg	acaggtatcc	ggtaagcggc	4440
agggtcggaa	caggagagcg	cacgagggag	cttcacgggg	gaaacgcctg	gtatctttat	4500
agtcctgtcg	ggtttcgcca	cctctgactt	gagcgtcgat	ttttgtgatg	ctcgtcaggg	4560
ggcgcgagcc	tatggaaaaa	cgccagcaac	gcggcctttt	tacggttcct	ggccttttgc	4620
tggccttttg	ctcacatggt	ctttcctgcg	ttatccctcg	attctgtgga	taaccgtatt	4680
accgccatgc	at					4692

&lt;210&gt; 30

&lt;211&gt; 4257

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pPur plasmid from Clontech

&lt;400&gt; 30

ctgtggaatg	tgtgtcagtt	agggtgtgga	aagtccccag	gctccccagc	aggcagaagt	60
atgcaaagca	tgcattctcaa	ttagtcagca	accaggtgtg	gaaagtcccc	agggtcccca	120
gcaggcagaa	gtatgcaaaag	catgcatctc	aattagtcag	caaccatagt	cccgcccccta	180
actccgcccc	tccccgccct	aactccgccc	agttccgccc	attctccgcc	ccatggctga	240
ctaatttttt	ttattttatgc	agaggccgag	gccgcctcgg	cctctgagct	attccagaag	300
tagtgaggag	gcttttttgg	aggcctaggc	ttttgcaaaa	agcttgcatg	cctgcaggtc	360
ggcggccacg	accggtgccc	ccaccatccc	ctgaccacag	ccccctgacc	ctcacaagga	420
gacgaccttc	catgaccgag	tacaagcccc	cgggtgcgct	cgccacccgc	gacgacgtcc	480
cccgggcccgt	acgcaccctc	gccgcccgtg	tcgcccacta	ccccgccacg	cgccacaccc	540
tcgacccgga	ccgccacatc	gagcgggtca	ccgagctgca	agaactcttc	ctcacgcgcg	600
tcgggctcga	catcggcaag	gtgtgggtcg	cggacgacgg	cgccgcgggt	gcgggtctgga	660
ccacgcggga	gagcgtcgaa	gcggggggcg	tgttcgccga	gatcggcccg	cgcattggccg	720
agttgagcgg	ttcccggtcg	gccgcgcagc	aacagatgga	aggcctcctg	gcgcccgcacc	780
ggcccaagga	gcccgcgtgg	ttcctggcca	ccgtcggcgt	ctcgcccgcac	caccagggga	840
agggtctggg	cagcgcgcgt	gtgctccccg	gagtgaggcg	ggccgagcgc	gccgggggtgc	900
ccgccttcct	ggagacctcc	gcgccccgca	acctcccttc	ctacgagcgg	ctcggcttca	960
ccgtcacccg	cgacgtcgag	gtgcccgaag	gaccgcgcac	ctggtgcatg	acccgcgaagc	1020
ccggtgctcg	acgcccgcgc	cacgacccgc	agcgcgccag	cgaaaggagc	gcacaccccc	1080
.atggctccga	ccgaagccga	cccggggcgg	cccgcgcacc	ccgcacccgc	ccccgaggcc	1140
caccgactct	agaggatcat	aatcagccat	accacatttg	tagaggtttt	acttgcttta	1200
aaaaacctcc	cacacctccc	cctgaacctg	aaacataaaa	tgaatgcaat	tggttgttgt	1260
aacttgttta	ttgcagctta	taatggttac	aaataaagca	atagcatcac	aaatttcaca	1320
aataaagcat	ttttttcact	gcattctagt	tgtgtgttgt	ccaaactcat	caatgtatct	1380
tatcatgtct	ggatccccag	gaagctcctc	tgtgtcctca	taaaccctaa	cctcctctac	1440
ttgagaggac	attccaatca	taggctgccc	atccaccctc	tgtgtcctcc	tggttaattag	1500
gtcacttaac	aaaaaggaaa	ttgggtaggg	gtttttcaca	gaccgctttc	taagggtaat	1560
tttaaaatat	ctgggaagtc	ccttcacatg	ctgtgttcca	gaagtgttgg	taaacagccc	1620
acaaatgtca	acagcagaaa	catacaagct	gtcagctttg	cacaagggcc	caacaccctg	1680
ctcatcaaga	agcactgtgg	ttgctgtggt	agtaatgtgc	aaaacaggag	gcacattttc	1740
cccacctgtg	taggttccaa	aatatctagt	gttttcatct	ttacttggat	caggaaccca	1800
gcactccact	ggataagcat	tatccttatc	caaaacagcc	ttgtggtcag	tggtcatctg	1860
ctgactgtca	actgtagcat	tttttggggg	tacagtttga	gcaggatatt	tggtcctgta	1920
gtttgtctaa	acaccctgca	gctccaaagg	ttccccacca	acagcaaaaa	aatgaaaatt	1980
tgacccttga	atgggttttc	cagcaccatt	ttcatgagtt	ttttgtgtcc	ctgaatgcaa	2040
gtttaacata	gcagttaccc	caataacctc	agttttaaca	gtaacagctt	cccacatcaa	2100
aatattttcca	caggttaagt	cctcatttaa	attaggcaaa	ggaattcttg	aagacgaaag	2160
ggcctcgtga	tacgcctatt	tttataggtt	aatgtcatga	taataatggg	ttcttagacg	2220
tcagggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	tttgtttatt	tttctaaata	2280
cattcaataa	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	ataatattga	2340
aaaagggaaga	tgtatagtat	tcaacatttc	cgtgtcgccc	ttattccctt	ttttcgggca	2400
ttttgccttc	ctgtttttgc	tcacccagaa	acgctggtga	aagtaaaaga	tgctgaagat	2460
cagttgggtg	ccgaagtggg	ttacatcgaa	ctggatctca	acagcggtaa	gatccttgag	2520
agtttttcg	cctgaagaacg	ttttccaatg	atgagcactt	ttaaagtctc	gctatgtggg	2580
gcgggtattat	cccgtgttga	cgccggggcaa	gagcaactcg	gtcgcgcgat	acactattct	2640
cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	tggcatgaca	2700
tgaagagaat	tatgcagtcg	tgccataacc	atgagtgata	acactgcggc	caacttactt	2760

-17-

ctgacaacga	tccgaggacc	gaaggagcta	accgcttttt	tgcacaacat	gggggatcat	2820
gtaactcgcc	ttgatcggtg	ggaaccggag	ctgaatgaag	ccataccaaa	cgacgagcgt	2880
gacaccacga	tgcctgcagc	aatggcaaca	acgttgcgca	aactattaac	tggcgaacta	2940
cttactctag	cttcccggca	acaattaata	gactggatgg	agggcgataa	agttgcagga	3000
ccacttctgc	gctcgccctt	tccggctggc	tggtttattg	ctgataaatc	tggagccggg	3060
gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggtaagcc	ctcccgtatc	3120
gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	acagatcgct	3180
gagatagggt	cctcactgat	taagcattgg	taactgtcag	accaagttta	ctcatatata	3240
ctttagattg	atttaaaact	tcatttttaa	tttaaaagga	tctaggtgaa	gacccctttt	3300
gataatctca	tgacccaaa	cccttaacgt	gagttttcgt	tccactgagc	gtcagacccc	3360
gtagaaaaga	tcaaaggatc	ttcttgagat	cccttttttc	tgcgcgtaat	ctgctgcttg	3420
caaacaaaaa	aaccacccgt	accagcgggt	gtttgtttgc	cggatcaaga	gctaccaact	3480
ctttttccga	aggttaactg	cttcagcaga	gcgcagatac	caaataactgt	ccttctagt	3540
tagccgtagt	taggccacca	cttcaagaac	tctgtagcac	cgcctacata	cctcgctctg	3600
ctaactcgtg	taccagtggc	tgctgccagt	ggcgataagt	cgtgtcttac	cgggttggac	3660
tcaagacgat	agttaccgga	taaggcgcag	cggtcgggct	gaacgggggg	ttcgtgcaca	3720
cagcccagct	tggagcgaac	gacctacacc	gaactgagat	acctacagcg	tgagctatga	3780
gaaagcgcca	cgcttcccga	agggagaaa	gcggacaggt	atccggtaag	cggcagggtc	3840
ggaacaggag	agcgcacgag	ggagcttcca	gggggaaacg	cctgggtatct	ttatagtcct	3900
gtcgggtttc	gccacctctg	acttgagcgt	cgatttttgt	gatgctcgtc	agggggggcg	3960
agcctatgga	aaaacgccag	caacgcggcc	tttttacggg	tcctggcctt	ttgctggcct	4020
tttgctcaca	tggtctttcc	tgcggttatcc	cctgattctg	tggataaccg	tattaccgcc	4080
tttgagttag	ctgataccgc	tcgccgcagc	cgaacgaccg	agcgcagcga	gtcagtgagc	4140
gaggaagcgg	aagagcgcc	gatgcggtat	ttctcctta	cgcactctgt	cggattttca	4200
caccgcata	ggtgcactct	cagtacaatc	tgctctgatg	ccgcatagtt	aagccag	4257

&lt;210&gt; 31

&lt;211&gt; 8136

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pWE15 cosmid vector

&lt;300&gt;

&lt;308&gt; GenBank X65279

&lt;309&gt; 1995-04-14

&lt;400&gt; 31

ctatagttag	tcgtattatg	cggccgcgaa	ttcttgaaga	cgaaggggcc	tcgtgatacg	60
cctattttta	taggttaatg	tcattgataat	aatgggttct	tagacgtcag	gtggcacttt	120
tcggggaaat	tgccgcggaa	cccctatttg	tttatttttc	taaatacatt	caaatatgta	180
tcgcgtcatg	agacaataac	cctgataaat	gcttcaataa	tattgaaaaa	ggaagagtat	240
gagtattcaa	cattttccgtg	tcgcccttat	tccttttttt	gcggcatttt	gcttccgtgt	300
tttgctcacc	cagaaaacgct	ggtgaaaagt	aaagatgctg	aagatcagtt	gggtgcacga	360
gtgggttaca	tcgaactgga	tctcaacagc	ggtaagatcc	ttgagagtgt	tcgccccgaa	420
gaacgttttc	caatgatgag	cactttttaa	gttctgctat	gtggcgcggt	attatcccgt	480
ggtgacgccg	ggcaagagca	actcggtcgc	cgcatacact	attctcagaa	tgacttgggt	540
gagtactcac	cagtcacaga	aaagcatctt	acggatggca	tgacagtaag	agaattatgc	600
agtgtgcaca	taaccatgag	tgataaacact	gcggccaact	tacttctgac	aacgatcgga	660
ggaccgaagg	agctaaccgc	ttttttgcac	aacatggggg	atcatgtaac	tcgccttgat	720
cggtgggaac	cggagctgaa	tgaagccata	ccaaacgacg	agcgtgacac	cacgatgcct	780
gcagcaatgg	caacaacgtt	gcgcaaaacta	ttaactggcg	aactacttac	tctagcttcc	840
cggcaacaat	taatagactg	gatggaggcg	gataaagtgt	caggaccact	tctgcgctcg	900
gcccttccgg	ctggctgggt	tattgctgat	aaatctggag	ccggtgagcg	tgggtctcgc	960
ggtatcattg	cagcactggg	gccagatggt	aagccctccc	gtatcgtagt	tatctacacg	1020
acggggagtc	aggcaactat	ggatgaacga	aatagacaga	tcgctgagat	aggtgcctca	1080
ctgattaaag	attgttaact	gtcagaccaa	gtttactcat	atatacttta	gattgattta	1140
aaacttcatt	tttaatttaa	aaggatctag	gtgaagatcc	tttttgataa	tctcatgacc	1200
aaaatccctt	aacgtgagtt	ttcgttccac	tgagcgtcag	accccgtaga	aaagatcaaa	1260
ggatcctctt	gagatccttt	ttttctcgcg	gtaatctgct	gcttgcaaac	aaaaaaacca	1320
ccgctaccag	cggtgggtttg	tttgccggat	caagagctac	caactccttt	tccgaaggta	1380
actggcttca	gcagagcgca	gataccaaat	actgtccttc	tagtgtagcc	gtagttagcc	1440
caccacttca	agcaactctgt	agcaccgcct	acatacctcg	ctctgcta	cctgttacca	1500
gtggctgctg	ccagtgccga	taagtcgtgt	cctaccgggt	tggactcaag	acgatagtta	1560
cgggataagg	cgcagcggtc	gggctgaacg	gggggttcgt	gcacacagcc	cagcttggag	1620
cgaacgacct	acaccgaact	gagataccta	cagcgtgagc	tatgagaaag	cggcacgctt	1680



ccgaagggag	aaagggcgac	aggtatccgg	taagcggcag	ggcgggaaca	ggagagcgca	1740
cgagggagct	tccaggggga	aacgcctggg	atcttttatag	tcctgtcggg	gtttcgccac	1800
ctctgacttg	agcgtcgatt	tttgtgatgc	tcgtcagggg	ggcggagcct	atggaaaaac	1860
gccagcaacg	cggccttttt	acggttcctg	gcctttttgct	ggccttttgc	tcacatgttc	1920
tttctcgctg	tatcccctga	ttctgtggat	aacggtatta	ccgcctttga	gtgagctgat	1980
accgctcgcc	gcagccgaac	gaccgagcgc	agcaggtcag	tgagcgagga	agcgggaagag	2040
cgctgacttc	cgcgtttcca	gacttttacga	aacacggaaa	ccgaagacca	ttcatgttgt	2100
tgctcaggtc	gcagacgttt	tgacgcagca	gtcgcttcac	gttcgctcgc	gtatcggtga	2160
ttcattctgc	taaccagtaa	ggcaaccccc	ccagcctagc	cgggtcctca	acgacaggag	2220
cacgatcatg	cgcacccgtc	agatccagac	atgataagat	acattgatga	gtttggacaa	2280
accacaacta	gaatgcagtg	aaaaaaatgc	tttatttgtg	aaatttgtga	tgctattgct	2340
ttatttgtaa	ccattataag	ctgcaataaa	caagttaaca	acaacaattg	cattcatttt	2400
atgtttcagg	ttcaggggga	gggtgtgggag	gttttttaaa	gcaagtaaaa	cctctacaaa	2460
ttgtggtagg	ctgattatga	tctctagtca	aggcactata	catcaaata	tccttattaa	2520
cccctttaca	aattaaaaag	ctaaagggtac	acaatttttg	agcatagtta	ttaatagcag	2580
acactctatg	cctgtgtgga	gtaagaaaaa	acagtatgtt	atgattataa	ctggttatgcc	2640
tacttataaa	ggttacagaa	tatttttcca	taattttctt	gtatagcagt	gcagcttttt	2700
cctttgtggg	gtaaatagca	aagcaagcaa	gagttctatt	actaaacaca	gcatgactca	2760
aaaaacttag	caattctgaa	ggaaagtcct	tggggctctc	tacctttctc	ttcttttttg	2820
gaggagtaga	atgtttgagag	tcagcagtag	cctcatcatc	actagatggc	attttctctg	2880
agcaaaacag	gttttcctca	ttaaaggcat	tcaccactg	ctccattca	tcagttccat	2940
aggttgggaat	ctaaaataca	caaacaatta	gaatcagtag	tttaacacat	tatacactta	3000
aaaattttat	attttacctta	gagcttttaa	tctctgtagg	tagtttgtcc	aattatgtca	3060
caccacagaa	gtaaggttcc	ttcacaaaga	tccggaccac	agcggccatc	gtgcctcccc	3120
actcctgcag	ttcgggggca	tggatgcgcg	gatagccgct	gctgggttcc	tggatgcoga	3180
cggatttgca	ctgccggtag	aactcgcgag	gtcgtccagc	ctcaggcagc	agctgaacca	3240
actcgcgagg	ggatcgagcc	cgggggtggc	gaagaactcc	agcatgagat	ccccgcctg	3300
gaggatcatc	cagccggcgt	cccggaaaac	gattccgaag	cccaaccttt	catagaaggc	3360
ggcgggtgaa	tcgaaatctc	gtgatggcag	gttggggcgt	gcttgggtcgg	tcatttcogaa	3420
cccagagtc	ccgctcagaa	gaactcgtoa	agaaggcgat	agaaggcgat	gcgctgcgaa	3480
tcgggagcgg	cgataccgta	aagcacgagg	aagcggtcag	cccattcgcc	gccaagctct	3540
tcagcaatat	cacgggtagc	caacgctatg	tctgatagc	ggtccgccac	accagccggg	3600
ccacacagaa	tgaatccaga	aaagcggcca	ttttccacca	tgatattcgg	caagcaggga	3660
tcgccatggg	tcacgacgag	atcctcgccg	tcgggatgcg	cgccttgagc	ctggcgaaaca	3720
gttcggctgg	cgcgagcccc	tgatgctctt	cgtccagatc	atcctgatcg	acaagaccgg	3780
cttccatccg	agtaacgtgt	cgtccgatgc	gatgtttcgc	ttgggtggtcg	aatgggcagg	3840
tagccggatc	aagcgtatgc	agccgcgcga	ttgcatcagc	catgatggat	actttctcgg	3900
caggagcaag	gtgagatgac	aggagatcct	gccccggcac	ttcgcccaat	agcagccagt	3960
cccttccgc	ttcagtgaca	acgtcgagca	cagctggcca	aggaacgcgc	gtcgtggcca	4020
gccacgatag	ccgcgctgcc	tcgtcctgca	gttcattcag	ggcaccggac	aggtcgggtct	4080
tgacaaaaag	aaccggggcg	ccctgcgctg	acagccggaa	cacggcggca	tcagagcagc	4140
cgatatgtct	ttgtgccagt	tcatagcctc	atagcctctc	cacccaagcg	gccggagaac	4200
ctgcgtgcaa	tccatcttgt	tcaatcatgc	gaaacgatcc	tcattcctgt	tcttgatcag	4260
atcttgatcc	cctgcgccat	cagatccttg	gcggaagaaa	agccatccag	tttactttgc	4320
agggcttccc	aaccttacca	gagggcgccc	cagctggcaa	ttccgggttcg	cttgctgtcc	4380
ataaaaaccgc	ccagttctagc	tatcgccatg	taagccact	gcaagctacc	tgctttctct	4440
ttgcgcttgc	gttttcctct	gtccagatag	cccagtagct	gacattcatc	cggggctcagc	4500
accgtttctg	cggactggct	ttctacgtgt	tccgcttctt	ttagcagccc	ttgcgcctcg	4560
agtgtcttgc	gcagcgtgaa	agcttttttg	aaaagcctag	gcctccaaaa	aagcctctct	4620
actacttctg	gaatagctca	gaggccgagg	cggcctaaat	aaaaaaaatt	agtcagccat	4680
ggggcggaga	atgggcggaa	ctgggcggag	ttagggcgcg	gatgggcgga	gttaggggag	4740
ggactatggg	tgctgactaa	ttgagatgca	tgctttgcat	acttctgcct	gctggggagc	4800
ctggggactt	tccacacctg	gttgctgact	aattgagatg	catgctttgc	atacttctgc	4860
ctgctgggga	gcctggggac	tttccacacc	ctaactgaca	cacattccac	agccggatct	4920
gcaggaccca	agcctgcccg	agatgcgcgc	cgtgcggctg	ctggagatgg	cggacgcgat	4980
ggatatgttc	tgccaagggg	tggtttgcgc	attcacagtt	ctccgcaaga	attgatggag	5040
tccaattctt	ggagtggtga	atccgttagc	gaggtgcgcg	cggcttccat	tcaggtccgc	5100
gtggcccgcc	ttcatgacac	gcgacgcaac	gcggggaggc	agacaaggta	tagggcgggc	5160
cctacaatcc	atgccaaacc	gttccatgtg	ctcgccgagg	cgcataaatc	gccgtgacga	5220
ccagcgggtc	aatgatcgaa	gttaggtggg	taagagccgc	gagcgatcct	tgaagctgtc	5280
cctgatgggtc	gtcatctacc	tgcttgagca	gcattggcctg	caacgcggca	tcccgatgcc	5340
gccggaagcg	agaagaatca	taatggggga	ggccatccag	cctcgcgctc	cgaacgccag	5400
caagacgtag	ccagcgcggt	cgggcggcca	tgccggcgag	aatggcctgc	ttctcgcca	5460
aacglttggt	ggcgggacca	gtgacgaagg	cttgagcgag	ggcgtgcaag	attccgaata	5520
ccgcaagcga	caggccgatc	atcgtcgccg	tccagcgaaa	gcggctcctc	ccgaaaaatga	5580
cccagagcgc	tgccggcacc	tgctctacga	gttgcatgat	aaagaagaca	gtcataagtg	5640
cggcgacgat	agtcatgccc	cgcgcccacc	ggaaggagct	gactgggttg	aaggctctca	5700



-19-

agggcatcg	tgcagcgtct	cccttatg	actcctgcat	taggaagcag	cccagtagta	5760
gggtgagcc	ggtgagcacc	gccgcgcg	ggaatgggtg	atgcaaggag	atggcgccca	5820
acagtcctcc	ggccacgggc	ctgccaccat	acccacgcgc	aaacaagcgc	tcatgagccc	5880
gaagtggcga	gcccgatctt	ccccatcggt	gatgtcggcg	atatagggcg	cagcaaccgc	5940
acctgtggcg	ccggtgatgc	cggccacgat	gcgtccggcg	tagaggatct	tggcagtcac	6000
agcatgcgca	tatccatgct	tgcaccatgc	gctcacaag	taggtgaatg	cgcaatgtag	6060
taccacacatc	gtcatcgctt	tccactgctc	tcgcgaataa	agatggaaaa	tcaatctcat	6120
ggtaatatg	catgaaaatc	cttgatttca	taaatcctcc	aggtagctat	atgcaaattg	6180
aaacaaaaga	gatgggtgatc	tttctaagag	atgatggaat	ctcccttcag	tatcccgatg	6240
gtcaatgcgc	tggatatggg	atagatggga	atatgtgat	ttttatggga	cagagttg	6300
aactgttccc	aactaaaatc	attttgcacg	atcagcgac	tacgaacttt	acccacaaat	6360
agtcaggtaa	tgaatcctga	tataaagaca	gggtgataaa	tcagtcttct	acgcgcacgc	6420
cacgcgcaca	ccgtagaaaag	tctttcagtt	gtgagcctgc	gcaaaccggt	aactttcg	6480
ggctttgctg	tcgcacaggc	tcacgtctaa	aaggaaataa	atcatgggtc	ataaaattat	6540
cacgttgtcc	ggcgcgccga	cggatgttct	gtatgcgctg	tttttcgctg	gcgcgttgct	6600
gtctgggtgat	ctgccttcta	aatctggcac	agccgaattg	cgcgagcttg	gttttgcgtg	6660
aaccagacac	acagcaactg	aataccagaa	agaaaaatc	tttaccttct	tgacatcaga	6720
agggcagaaa	tttgccgttg	aacacctggt	caataacgcg	tttgggtgag	agcaatattg	6780
cgcttcgatg	acgcttggcg	ttgagattga	tacctctgct	gcacaaaagg	caatcgacga	6840
gctggaccag	cgcattcgtg	acaccgtctc	cttcgaactt	attcgcaatg	gagtgctcatt	6900
catcaaggac	gccgctatcg	caaatggtgc	tatccacgca	gcggaactcg	aaacacctca	6960
gccggtgacc	aatatctaca	acatcagcct	tggtatccag	cgatgatgag	cagcgcagaa	7020
caaggttaacc	gtcagtgccg	ataagttcaa	agttaaacct	ggtgttgata	ccaacattga	7080
aacgttgatc	gaaaaacgcg	tgaaaaacgc	tgctgaatgt	gcggcgctgg	atgtcacaaa	7140
gcaaatggca	gcagacaaga	aagcgatgga	tgaactggct	tcctatgtcc	gcacggccat	7200
catgatggaa	tgtttccccc	gtgggtgttat	ctggcagcag	tgccgtcgat	agtatgcaat	7260
tgataattat	tatcatttgc	gggtcctttc	cgcgatccg	ccttgttacg	ggcgccgcac	7320
ctcgcgggtt	ttcgctat	atgaaaat	tccggtttta	ggcgtttccg	ttcttcttcg	7380
tcataactta	atgtttttat	ttaaaatacc	ctctgaaaag	aaaggaaaacg	acaggtgctg	7440
aaagcagcgt	ttttggcctc	tgctgtttcc	tttctctggt	tttgtccgtg	gaatgaacaa	7500
tggaagtcaa	caaaaagcag	ctggctgaca	ttttcggtgc	gagtatccgt	accattcaga	7560
actggcagga	acagggaatg	ccgttctg	gaggcggtgg	caagggtaat	gaggtgcttt	7620
atgactctgc	cgccgtcata	aaatggtatg	ccgaaaggga	tgctgaaatt	gagaacgaaa	7680
agctgcgccg	ggaggttgaa	gaactgcggc	aggccagcga	ggcagatcca	caggacgggt	7740
gtggctgcga	tgatcgcgta	gtcgatagt	gctccaagta	gcgaagcgag	caggactggg	7800
cggcgcaaaa	cggtcgggac	agtgtccga	gaacgggtgc	gcatagaaat	tgcatcaacg	7860
catatagcgc	tagcagcacg	ccatagtgc	tgccgatgct	gtcggaatgg	acgatatccc	7920
gcaagaggcc	cggcagtagc	ggcataacca	agcctatg	tacagcatcc	agggtagcgg	7980
tgccgaggat	gacgatgagc	gcattgttag	atctcataca	cggtgcctga	ctgcgttagc	8040
aatttaactg	tgataaaacta	ccgcattaaa	gcttatcgat	gataagcgg	caaacatgag	8100
aattcgccgc	cgcaattaac	cctcactaaa	ggatcc			8136

<210> 32  
 <211> 2713  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pNEB193 plasmid

<400> 32						
tcgcgcgttt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccc	gagacgggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accatagcgc	gtgtgaaata	ccgcacagat	gcgtaaggag	aaaataccgc	atcaggcgcc	240
attcgccatt	caggctgcgc	aactgttggg	aaggcgatc	ggtgcggggc	tcttcgctat	300
tacgccagct	ggcgaaaagg	ggatgtgctg	caaggcgatt	aagttgggta	acgccagggt	360
tttcccagtc	acgacgttgt	aaaacgacgg	ccagtgaaat	cgagctcggt	acccgggggc	420
gcgcgggatc	cttaatttaag	tctagagtcg	actgtttaaa	cctgcaggca	tgcaagcttg	480
gcgtaatcat	ggatcatagct	gtttcctgtg	tgaaattgtt	atccgctcac	aattccacac	540
aacatacag	ccggaagcat	aaagtgtaaa	ccctgggggtg	cctaagtgtg	gagctaaactc	600
acattaattg	cggtgcgctc	actgcccgc	ttccagtcgc	gaaacctgtc	gtgccagctg	660
cattaatgaa	tcggccaacg	cgcggggaga	ggcggtttgc	gtattgggcg	ctcttcgct	720
tcctcgctca	tcgctcgctc	gcgctcgctc	gttcggctgc	ggcgagcggt	atcagctcac	780
tcaaaaggcg	taatacgggt	atccacagaa	tcaggggata	acgcaggaaa	gaacatgtga	840
gcaaaaggcc	agcaaaaggc	caggaaccgt	aaaaaggccg	cggtgctggc	gtttttccat	900
aggctccgc	cccctgacga	gcatacaaaa	aatcgacgct	caagtacag	gtggcgaaac	960

-20-

ccgacaggac	tataaagata	ccaggcggtt	ccccctggaa	gctccctcgt	gcgctctcct	1020
gttccgaacc	tgccgcttac	cggataacct	tccgcctttc	tcccttcggg	aagcgtggcg	1080
ctttctcata	gctcacgctg	taggtatctc	agttcgggtg	aggctcgttcg	ctccaagctg	1140
ggctgtgtgc	acgaaccccc	cgttcagccc	gaccgctgcg	ccttatccgg	taactatcgt	1200
cttgagtcca	acccggttaag	acacgactta	tccgccactgg	cagcagccac	tggtaacagg	1260
attagcagag	cgaggtatgt	aggcggtgct	acagagttct	tgaagtgggtg	gcctaactac	1320
ggctacacta	gaaggacagt	atcttggtatc	tgcgctctgc	tgaagccagt	taccttcgga	1380
aaaagagttg	gtagctcttg	atccggcaaa	caaaccaccg	ctggtagcgg	tggttttttt	1440
gtttgcaagc	agcagattac	gcgcaaaaaa	aaaggatctc	aagaagatcc	tttgatcttt	1500
tctacggggg	ctgacgctca	gtggaacgaa	aactcacgtt	aagggatttt	ggtcatgaga	1560
ttatcaaaaa	ggatcttcac	ctagatcctt	ttaaattaaa	aatgaagttt	taaatcaatc	1620
taaagtatat	atgagtaaac	ttggtctgac	agttaccaat	gottaatcag	tgaggcacct	1680
atctcagcga	tctgtctatt	tcggtcatcc	atagttgcct	gactccccgt	cgtgtagata	1740
actacgatac	gggagggcctt	accatctggc	cccagtgtcg	caatgatacc	gcgagacca	1800
cgctcaccgg	ctccagattt	atcagcaata	aaccagccag	ccggaagggc	cgagcgcaga	1860
agtggtcctg	caactttatc	cgcctccatc	cagtctatta	attggtgccc	ggaagctaga	1920
gtaagttagt	gccagtttaa	tagtttgctg	aacgttggtg	ccattgctac	agggatcggtg	1980
gtgtcacgct	cgctcgtttg	tatggcttca	ttcagctccg	gttcccaacg	atcaaggcga	2040
gttacatgat	cccccatgtt	gtgcaaaaaa	gcggttagct	ccttcgggtc	tccgatcgtt	2100
gtcagaagta	agttggccgc	agtgttatca	ctcatgggta	tggcagcact	gcataattct	2160
cttactgtca	tgccatccgt	aagatgcttt	tctgtgactg	gtgagtactc	aaccaagtca	2220
ttctgagaat	agtgtatgcg	gcgaccgagt	tgtctctgce	cggcgtcaat	acgggataat	2280
accgcgccac	atagcagaac	tttaaaaagt	ctcatcattg	gaaaacgttc	ttcggggcga	2340
aaactctcaa	ggatcttacc	gctgttgaga	tccagttcga	tgtaaccac	tcgtgcaccc	2400
aactgatctt	cagcatcttt	tactttcacc	agcgtttctg	ggtgagcaaa	aacaggaagg	2460
caaaatgccg	caaaaaaggg	aataagggcg	acacggaaat	ggtgaatact	catactcttc	2520
ctttttcaat	attattgaag	catttatcag	ggttattgtc	tcattgagcg	atacatattt	2580
gaatgtatct	agaaaaataa	acaaataggg	gttccgcgca	catttccccg	aaaagtgcca	2640
cctgacgtct	aagaaaccat	tattatcatg	acattaacct	ataaaaaatg	gcgtatcacg	2700
aggcccttcc	gtc					2713

<210> 33  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> attP

<400> 33  
 cagctttttt atactaagtt g 21

<210> 34  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> attB

<400> 34  
 ctgctttttt atactaactt g 21

<210> 35  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> attL

<400> 35  
 ctgctttttt atactaagtt g 21

<210> 36  
 <211> 21  
 <212> DNA

-21-

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; attR

&lt;400&gt; 36

cagctttttt atactaactt g

21

&lt;210&gt; 37

&lt;211&gt; 1071

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Integrase E174R

&lt;221&gt; CDS

&lt;222&gt; (1)...(1071)

&lt;223&gt; Nucleotide sequence encoding Integrase E147R

&lt;400&gt; 37

atg gga aga agg cga agt cat gag cgc cgg gat tta ccc cct aac ctt	48
Met Gly Arg Arg Arg Ser His Glu Arg Arg Asp Leu Pro Pro Asn Leu	
1 5 10 15	
tat ata aga aac aat gga tat tac tgc tac agg gac cca agg acg ggt	96
Tyr Ile Arg Asn Asn Gly Tyr Tyr Cys Tyr Arg Asp Pro Arg Thr Gly	
20 25 30	
aaa gag ttt gga tta ggc aga gac agg cga atc gca atc act gaa gct	144
Lys Glu Phe Gly Leu Gly Arg Asp Arg Arg Ile Ala Ile Thr Glu Ala	
35 40 45	
ata cag gcc aac att gag tta ttt tca gga cac aaa cac aag cct ctg	192
Ile Gln Ala Asn Ile Glu Leu Phe Ser Gly His Lys His Lys Pro Leu	
50 55 60	
aca gcg aga atc aac agt gat aat tcc gtt acg tta cat tca tgg ctt	240
Thr Ala Arg Ile Asn Ser Asp Asn Ser Val Thr Leu His Ser Trp Leu	
65 70 75 80	
gat cgc tac gaa aaa atc ctg gcc agc aga gga atc aag cag aag aca	288
Asp Arg Tyr Glu Lys Ile Leu Ala Ser Arg Gly Ile Lys Gln Lys Thr	
85 90 95	
ctc ata aat tac atg agc aaa att aaa gca ata agg agg ggt ctg cct	336
Leu Ile Asn Tyr Met Ser Lys Ile Lys Ala Ile Arg Arg Gly Leu Pro	
100 105 110	
gat gct cca ctt gaa gac atc acc aca aaa gaa att gcg gca atg ctc	384
Asp Ala Pro Leu Glu Asp Ile Thr Thr Lys Glu Ile Ala Ala Met Leu	
115 120 125	
aat gga tac ata gac gag ggc aag gcg gcg tca gcc aag tta atc aga	432
Asn Gly Tyr Ile Asp Glu Gly Lys Ala Ala Ser Ala Lys Leu Ile Arg	
130 135 140	
tca aca ctg agc gat gca ttc cga gag gca ata gct gaa ggc cat ata	480
Ser Thr Leu Ser Asp Ala Phe Arg Glu Ala Ile Ala Glu Gly His Ile	
145 150 155 160	
aca aca aac cat gtc gct gcc act cgc gca gca aaa tct aga gta agg	528
Thr Thr Asn His Val Ala Ala Thr Arg Ala Ala Lys Ser Arg Val Arg	
165 170 175	
aga tca aga ctt acg gct gac gaa tac ctg aaa att tat caa gca gca	576
Arg Ser Arg Leu Thr Ala Asp Glu Tyr Leu Lys Ile Tyr Gln Ala Ala	

-22-

180										185										190										
gaa	tca	tca	cca	tgt	tgg	ctc	aga	ctt	gca	atg	gaa	ctg	gct	gtt	gtt	624														
Glu	Ser	Ser	Pro	Cys	Trp	Leu	Arg	Leu	Ala	Met	Glu	Leu	Ala	Val	Val															
		195					200					205																		
acc	ggg	caa	cga	gtt	ggg	gat	tta	tgc	gaa	atg	aag	tgg	tct	gat	atc	672														
Thr	Gly	Gln	Arg	Val	Gly	Asp	Leu	Cys	Glu	Met	Lys	Trp	Ser	Asp	Ile															
	210					215					220																			
gta	gat	gga	tat	ctt	tat	gtc	gag	caa	agc	aaa	aca	ggc	gta	aaa	att	720														
Val	Asp	Gly	Tyr	Leu	Tyr	Val	Glu	Gln	Ser	Lys	Thr	Gly	Val	Lys	Ile															
	225				230					235					240															
gcc	atc	cca	aca	gca	ttg	cat	att	gat	gct	ctc	gga	ata	tca	atg	aag	768														
Ala	Ile	Pro	Thr	Ala	Leu	His	Ile	Asp	Ala	Leu	Gly	Ile	Ser	Met	Lys															
				245				250						255																
gaa	aca	ctt	gat	aaa	tgc	aaa	gag	att	ctt	ggc	gga	gaa	acc	ata	att	816														
Glu	Thr	Leu	Asp	Lys	Cys	Lys	Glu	Ile	Leu	Gly	Gly	Glu	Thr	Ile	Ile															
			260					265					270																	
gca	tct	act	cgt	cgc	gaa	ccg	ctt	tca	tcc	ggc	aca	gta	tca	agg	tat	864														
Ala	Ser	Thr	Arg	Arg	Glu	Pro	Leu	Ser	Ser	Gly	Thr	Val	Ser	Arg	Tyr															
		275					280					285																		
ttt	atg	cgc	gca	cga	aaa	gca	tca	ggg	ctt	tcc	ttc	gaa	ggg	gat	ccg	912														
Phe	Met	Arg	Ala	Arg	Lys	Ala	Ser	Gly	Leu	Ser	Phe	Glu	Gly	Asp	Pro															
	290					295					300																			
cct	acc	ttt	cac	gag	ttg	cgc	agt	ttg	tct	gca	aga	ctc	tat	gag	aag	960														
Pro	Thr	Phe	His	Glu	Leu	Arg	Ser	Leu	Ser	Ala	Arg	Leu	Tyr	Glu	Lys															
	305				310					315					320															
cag	ata	agc	gat	aag	ttt	gct	caa	cat	ctt	ctc	ggg	cat	aag	tcg	gac	1008														
Gln	Ile	Ser	Asp	Lys	Phe	Ala	Gln	His	Leu	Leu	Gly	His	Lys	Ser	Asp															
				325					330					335																
acc	atg	gca	tca	cag	tat	cgt	gat	gac	aga	ggc	agg	gag	tgg	gac	aaa	1056														
Thr	Met	Ala	Ser	Gln	Tyr	Arg	Asp	Asp	Arg	Gly	Arg	Glu	Trp	Asp	Lys															
			340					345					350																	
att	gaa	atc	aaa	taa												1071														
Ile	Glu	Ile	Lys	*																										
		355																												

&lt;210&gt; 38

&lt;211&gt; 356

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Integrase E147R

&lt;400&gt; 38

Met	Gly	Arg	Arg	Arg	Ser	His	Glu	Arg	Arg	Asp	Leu	Pro	Pro	Asn	Leu
1				5					10					15	
Tyr	Ile	Arg	Asn	Asn	Gly	Tyr	Tyr	Cys	Tyr	Arg	Asp	Pro	Arg	Thr	Gly
			20					25					30		
Lys	Glu	Phe	Gly	Leu	Gly	Arg	Asp	Arg	Arg	Ile	Ala	Ile	Thr	Glu	Ala
		35					40				45				
Ile	Gln	Ala	Asn	Ile	Glu	Leu	Phe	Ser	Gly	His	Lys	His	Lys	Pro	Leu
	50					55					60				
Thr	Ala	Arg	Ile	Asn	Ser	Asp	Asn	Ser	Val	Thr	Leu	His	Ser	Trp	Leu
	65				70					75				80	
Asp	Arg	Tyr	Glu	Lys	Ile	Leu	Ala	Ser	Arg	Gly	Ile	Lys	Gln	Lys	Thr

[illegible]

```
<210> 39
<211> 876
<212> DNA
<213> Discosoma species

<220>
<221> CDS
<222> (45)...(737)
<223> Nucleotide sequence encoding red fluorescent
      protein (FP593)
```

<300>  
<308> GenBank AF272711  
<309> 2000-09-26

<400> 39																
agtttcagcc agtgacaggg tgagctgccca ggtattctaa caag atg agt tgt tcc																56
Met Ser Cys Ser																
1																
aag aat gtg atc aag gag ttc atg agg ttc aag gtt cgt atg gaa gga																104
Lys Asn Val Ile Lys Glu Phe Met Arg Phe Lys Val Arg Met Glu Gly																20
5 10 15																
acg gtc aat ggg cac gag ttt gaa ata aaa ggc gaa ggt gaa ggg agg																152
Thr Val Asn Gly His Glu Phe Glu Ile Lys Gly Glu Gly Glu Gly Arg																35
25 30																
cct tac gaa ggt cac tgt tcc gta aag ctt atg gta acc aag ggt gga																200
Pro Tyr Glu Gly His Cys Ser Val Lys Leu Met Val Thr Lys Gly Gly																

-24-

40										45										50										
cct	ttg	cca	ttt	gct	ttt	gat	att	ttg	tca	cca	caa	ttt	cag	tat	gga	248														
Pro	Leu	Pro	Phe	Ala	Phe	Asp	Ile	Leu	Ser	Pro	Gln	Phe	Gln	Tyr	Gly															
		55					60					65																		
agc	aag	gta	tat	gtc	aaa	cac	cct	gcc	gac	ata	cca	gac	tat	aaa	aag	296														
Ser	Lys	Val	Tyr	Val	Lys	His	Pro	Ala	Asp	Ile	Pro	Asp	Tyr	Lys	Lys															
	70					75					80																			
ctg	tca	ttt	cct	gag	gga	ttt	aaa	tgg	gaa	agg	gtc	atg	aac	ttt	gaa	344														
Leu	Ser	Phe	Pro	Glu	Gly	Phe	Lys	Trp	Glu	Arg	Val	Met	Asn	Phe	Glu															
	85				90					95					100															
gac	ggg	ggc	gtg	gtt	act	gta	tcc	caa	gat	tcc	agt	ttg	aaa	gac	ggc	392														
Asp	Gly	Gly	Val	Val	Thr	Val	Ser	Gln	Asp	Ser	Ser	Leu	Lys	Asp	Gly															
				105				110						115																
tgt	ttc	atc	tac	gag	gtc	aag	ttc	att	ggg	gtg	aac	ttt	cct	tct	gat	440														
Cys	Phe	Ile	Tyr	Glu	Val	Lys	Phe	Ile	Gly	Val	Asn	Phe	Pro	Ser	Asp															
			120					125					130																	
gga	cct	gtt	atg	cag	agg	agg	aca	cgg	ggc	tgg	gaa	gcc	agc	tct	gag	488														
Gly	Pro	Val	Met	Gln	Arg	Arg	Thr	Arg	Gly	Trp	Glu	Ala	Ser	Ser	Glu															
		135					140					145																		
cgt	ttg	tat	cct	cgt	gat	ggg	gtg	ctg	aaa	gga	gac	atc	cat	atg	gct	536														
Arg	Leu	Tyr	Pro	Arg	Asp	Gly	Val	Leu	Lys	Gly	Asp	Ile	His	Met	Ala															
	150					155					160																			
ctg	agg	ctg	gaa	gga	ggc	ggc	cat	tac	ctc	gtt	gaa	ttc	aaa	agt	att	584														
Leu	Arg	Leu	Glu	Gly	Gly	Gly	His	Tyr	Leu	Val	Glu	Phe	Lys	Ser	Ile															
	165				170					175					180															
tac	atg	gta	aag	aag	cct	tca	gtg	cag	ttg	cca	ggc	tac	tat	tat	gtt	632														
Tyr	Met	Val	Lys	Lys	Pro	Ser	Val	Gln	Leu	Pro	Gly	Tyr	Tyr	Tyr	Val															
				185				190						195																
gac	tcc	aaa	ctg	gat	atg	acg	agc	cac	aac	gaa	gat	tac	aca	gtc	gtt	680														
Asp	Ser	Lys	Leu	Asp	Met	Thr	Ser	His	Asn	Glu	Asp	Tyr	Thr	Val	Val															
			200					205					210																	
gag	cag	tat	gaa	aaa	acc	cag	gga	cgc	cac	cat	ccg	ttc	att	aag	cct	728														
Glu	Gln	Tyr	Glu	Lys	Thr	Gln	Gly	Arg	His	His	Pro	Phe	Ile	Lys	Pro															
		215					220					225																		
ctg	cag	tga	actcggtctca	gtcatggatt	agcggtaatg	gccacaaaag										777														
Leu	Gln	*																												
	230																													
gcacgatgat	cgtttttttag	gaatgcagcc	aaaaattgaa	ggttatgaca	gtagaaatac	837																								
aagcaacagg	ctttgcttat	taaacatgta	attgaaaac			876																								

&lt;210&gt; 40

&lt;211&gt; 230

&lt;212&gt; PRT

&lt;213&gt; Discosoma species

&lt;400&gt; 40

Met	Ser	Cys	Ser	Lys	Asn	Val	Ile	Lys	Glu	Phe	Met	Arg	Phe	Lys	Val
1				5					10					15	
Arg	Met	Glu	Gly	Thr	Val	Asn	Gly	His	Glu	Phe	Glu	Ile	Lys	Gly	Glu
			20					25					30		
Gly	Glu	Gly	Arg	Pro	Tyr	Glu	Gly	His	Cys	Ser	Val	Lys	Leu	Met	Val
		35					40					45			
Thr	Lys	Gly	Gly	Pro	Leu	Pro	Phe	Ala	Phe	Asp	Ile	Leu	Ser	Pro	Gln
	50					55					60				

-25-

Phe	Gln	Tyr	Gly	Ser	Lys	Val	Tyr	Val	Lys	His	Pro	Ala	Asp	Ile	Pro
65					70					75					80
Asp	Tyr	Lys	Lys	Leu	Ser	Phe	Pro	Glu	Gly	Phe	Lys	Trp	Glu	Arg	Val
				85					90					95	
Met	Asn	Phe	Glu	Asp	Gly	Gly	Val	Val	Thr	Val	Ser	Gln	Asp	Ser	Ser
			100					105					110		
Leu	Lys	Asp	Gly	Cys	Phe	Ile	Tyr	Glu	Val	Lys	Phe	Ile	Gly	Val	Asn
			115				120					125			
Phe	Pro	Ser	Asp	Gly	Pro	Val	Met	Gln	Arg	Arg	Thr	Arg	Gly	Trp	Glu
	130					135					140				
Ala	Ser	Ser	Glu	Arg	Leu	Tyr	Pro	Arg	Asp	Gly	Val	Leu	Lys	Gly	Asp
145					150					155					160
Ile	His	Met	Ala	Leu	Arg	Leu	Glu	Gly	Gly	His	Tyr	Leu	Val	Glu	
				165				170					175		
Phe	Lys	Ser	Ile	Tyr	Met	Val	Lys	Lys	Pro	Ser	Val	Gln	Leu	Pro	Gly
			180					185					190		
Tyr	Tyr	Tyr	Val	Asp	Ser	Lys	Leu	Asp	Met	Thr	Ser	His	Asn	Glu	Asp
		195					200					205			
Tyr	Thr	Val	Val	Glu	Gln	Tyr	Glu	Lys	Thr	Gln	Gly	Arg	His	His	Pro
	210					215					220				
Phe	Ile	Lys	Pro	Leu	Gln										
225					230										

<210> 41  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> m-att;

<221> misc\_difference  
 <222> 18  
 <223> n is a or g or c or t/u

<400> 41  
 rkycwgcttt yktrtacnaa stsgb

25

<210> 42  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> m-attB;

<221> misc\_difference  
 <222> 18  
 <223> n is a or g or c or t/u

<400> 42  
 agccwgcttt yktrtacnaa ctsgb

25

<210> 43  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> m-attR

<221> misc\_difference  
 <222> 18  
 <223> n is a or g or c or t/u

-26-

<400> 43  
gttcagcttt cktrtacnaa ctsgb 25

<210> 44  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> m-attL

<221> misc\_difference  
<222> 18  
<223> n is a or g or c or t/u

<400> 44  
agccwgcttt cktrtacnaa gtsgb 25

<210> 45  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> m-attP1

<221> misc\_difference  
<222> 18  
<223> n is a or g or c or t/u

<400> 45  
gttcagcttt yktrtacnaa gtsgb 25

<210> 46  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attB1

<400> 46  
agcctgcttt tttgtacaaa cttgt 25

<210> 47  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attB2

<400> 47  
agcctgcttt cttgtacaaa cttgt 25

<210> 48  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attB3

<400> 48  
accagcttt cttgtacaaa cttgt 25

<210> 49



-27-

<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attR1

<400> 49  
gttcagcttt tttgtacaaa cttgt 25

<210> 50  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attR2

<400> 50  
gttcagcttt cttgtacaaa cttgt 25

<210> 51  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attR3

<400> 51  
gttcagcttt cttgtacaaa gttgg 25

<210> 52  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attL1

<400> 52  
agcctgcttt tttgtacaaa gttgg 25

<210> 53  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attL2

<400> 53  
agcctgcttt cttgtacaaa gttgg 25

<210> 54  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attL3

<400> 54  
accagcttt cttgtacaaa gttgg 25

<210> 55  
<211> 25

-28-

<212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> attP1  
  
 <400> 55  
 gttcagcttt tttgtacaaa gttgg 25  
  
 <210> 56  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> attP2,P3  
  
 <400> 56  
 gttcagcttt cttgtacaaa gttgg 25  
  
 <210> 57  
 <211> 34  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Lox P site  
  
 <400> 57  
 ataacttcgt ataatgtatg ctatacgaag ttat 34  
  
 <210> 58  
 <211> 1032  
 <212> DNA  
 <213> Escherichia coli  
  
 <220>  
 <221> CDS  
 <222> (1)...(1032)  
 <223> nucleotide sequence encoding Cre recombinase  
  
 <400> 58  
 atg tcc aat tta ctg acc gta cac caa aat ttg cct gca tta ccg gtc 48  
 Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val  
 1 5 10 15  
  
 gat gca acg agt gat gag gtt cgc aag aac ctg atg gac atg ttc agg 96  
 Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg  
 20 25 30  
  
 gat cgc cag gcg ttt tct gag cat acc tgg aaa atg ctt ctg tcc gtt 144  
 Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val  
 35 40 45  
  
 tgc cgg tcg tgg gcg gca tgg tgc aag ttg aat aac cgg aaa tgg ttt 192  
 Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe  
 50 55 60  
  
 ccc gca gaa cct gaa gat gtt cgc gat tat ctt cta tat ctt cag gcg 240  
 Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala  
 65 70 75 80  
  
 cgc ggt ctg gca gta aaa act atc cag caa cat ttg ggc cag cta aac 288  
 Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn  
 85 90 95  
  
 atg ctt cat cgt cgg tcc ggg ctg cca cga cca agt gac agc aat gct 336

-29-

Met	Leu	His	Arg	Arg	Ser	Gly	Leu	Pro	Arg	Pro	Ser	Asp	Ser	Asn	Ala		
			100					105					110				
gtt	tca	ctg	gtt	atg	cgg	cgg	atc	cga	aaa	gaa	aac	gtt	gat	gcc	ggc		384
Val	Ser	Leu	Val	Met	Arg	Arg	Ile	Arg	Lys	Glu	Asn	Val	Asp	Ala	Gly		
		115					120					125					
gaa	cgt	gca	aaa	cag	gct	cta	gcg	ttc	gaa	cgc	act	gat	ttc	gac	cag		432
Glu	Arg	Ala	Lys	Gln	Ala	Leu	Ala	Phe	Glu	Arg	Thr	Asp	Phe	Asp	Gln		
	130					135					140						
gtt	cgt	tca	ctc	atg	gaa	aat	agc	gat	cgc	tgc	cag	gat	ata	cgt	aat		480
Val	Arg	Ser	Leu	Met	Glu	Asn	Ser	Asp	Arg	Cys	Gln	Asp	Ile	Arg	Asn		
	145				150					155					160		
ctg	gca	ttt	ctg	ggg	att	gct	tat	aac	acc	ctg	tta	cgt	ata	gcc	gaa		528
Leu	Ala	Phe	Leu	Gly	Ile	Ala	Tyr	Asn	Thr	Leu	Leu	Arg	Ile	Ala	Glu		
				165					170					175			
att	gcc	agg	atc	agg	gtt	aaa	gat	atc	tca	cgt	act	gac	ggc	ggg	aga		576
Ile	Ala	Arg	Ile	Arg	Val	Lys	Asp	Ile	Ser	Arg	Thr	Asp	Gly	Gly	Arg		
			180					185					190				
atg	tta	atc	cat	att	ggc	aga	acg	aaa	acg	ctg	gtt	agc	acc	gca	ggc		624
Met	Leu	Ile	His	Ile	Gly	Arg	Thr	Lys	Thr	Leu	Val	Ser	Thr	Ala	Gly		
			195			200						205					
gta	gag	aag	gca	ctt	agc	ctg	ggg	gta	act	aaa	ctg	gtc	gag	cga	tgg		672
Val	Glu	Lys	Ala	Leu	Ser	Leu	Gly	Val	Thr	Lys	Leu	Val	Glu	Arg	Trp		
	210					215					220						
att	tcc	gtc	tct	ggc	gta	gct	gat	gat	ccg	aat	aac	tac	ctg	ttt	tgc		720
Ile	Ser	Val	Ser	Gly	Val	Ala	Asp	Asp	Pro	Asn	Asn	Tyr	Leu	Phe	Cys		
	225				230					235					240		
cgg	gtc	aga	aaa	aat	ggc	gtt	gcc	gcg	cca	tct	gcc	acc	agc	cag	cta		768
Arg	Val	Arg	Lys	Asn	Gly	Val	Ala	Ala	Pro	Ser	Ala	Thr	Ser	Gln	Leu		
				245					250					255			
tca	act	cgc	gcc	ctg	gaa	ggg	att	ttt	gaa	gca	act	cat	cga	ttg	att		816
Ser	Thr	Arg	Ala	Leu	Glu	Gly	Ile	Phe	Glu	Ala	Thr	His	Arg	Leu	Ile		
			260					265					270				
tac	ggc	gct	aag	gat	gac	tct	ggc	cag	aga	tac	ctg	gcc	tgg	tct	gga		864
Tyr	Gly	Ala	Lys	Asp	Asp	Ser	Gly	Gln	Arg	Tyr	Leu	Ala	Trp	Ser	Gly		
		275					280					285					
cac	agt	gcc	cgt	gtc	gga	gcc	gcg	cga	gat	atg	gcc	cgc	gct	gga	gtt		912
His	Ser	Ala	Arg	Val	Gly	Ala	Ala	Arg	Asp	Met	Ala	Arg	Ala	Gly	Val		
	290					295					300						
tca	ata	ccg	gag	atc	atg	caa	gct	ggc	ggc	tgg	acc	aat	gta	aat	att		960
Ser	Ile	Pro	Glu	Ile	Met	Gln	Ala	Gly	Gly	Trp	Thr	Asn	Val	Asn	Ile		
	305				310					315					320		
gtc	atg	aac	tat	atc	cgt	aac	ctg	gat	agt	gaa	aca	ggg	gca	atg	gtg		1008
Val	Met	Asn	Tyr	Ile	Arg	Asn	Leu	Asp	Ser	Glu	Thr	Gly	Ala	Met	Val		
				325					330					335			
cgc	ctg	ctg	gaa	gat	ggc	gat	tag										1032
Arg	Leu	Leu	Glu	Asp	Gly	Asp	*										
			340														

<210> 59  
 <211> 343  
 <212> PRT

-30-

<213> *Escherichia coli*

&lt;400&gt; 59

```

Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val
1      5      10      15
Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg
      20      25      30
Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val
      35      40      45
Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe
      50      55      60
Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala
      65      70      75      80
Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn
      85      90      95
Met Leu His Arg Arg Ser Gly Leu Pro Arg Pro Ser Asp Ser Asn Ala
      100      105      110
Val Ser Leu Val Met Arg Arg Ile Arg Lys Glu Asn Val Asp Ala Gly
      115      120      125
Glu Arg Ala Lys Gln Ala Leu Ala Phe Glu Arg Thr Asp Phe Asp Gln
      130      135      140
Val Arg Ser Leu Met Glu Asn Ser Asp Arg Cys Gln Asp Ile Arg Asn
      145      150      155      160
Leu Ala Phe Leu Gly Ile Ala Tyr Asn Thr Leu Leu Arg Ile Ala Glu
      165      170      175
Ile Ala Arg Ile Arg Val Lys Asp Ile Ser Arg Thr Asp Gly Gly Arg
      180      185      190
Met Leu Ile His Ile Gly Arg Thr Lys Thr Leu Val Ser Thr Ala Gly
      195      200      205
Val Glu Lys Ala Leu Ser Leu Gly Val Thr Lys Leu Val Glu Arg Trp
      210      215      220
Ile Ser Val Ser Gly Val Ala Asp Asp Pro Asn Asn Tyr Leu Phe Cys
      225      230      235      240
Arg Val Arg Lys Asn Gly Val Ala Ala Pro Ser Ala Thr Ser Gln Leu
      245      250      255
Ser Thr Arg Ala Leu Glu Gly Ile Phe Glu Ala Thr His Arg Leu Ile
      260      265      270
Tyr Gly Ala Lys Asp Asp Ser Gly Gln Arg Tyr Leu Ala Trp Ser Gly
      275      280      285
His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val
      290      295      300
Ser Ile Pro Glu Ile Met Gln Ala Gly Gly Trp Thr Asn Val Asn Ile
      305      310      315      320
Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Thr Gly Ala Met Val
      325      330      335
Arg Leu Leu Glu Asp Gly Asp
      340

```

&lt;210&gt; 60

&lt;211&gt; 1272

&lt;212&gt; DNA

<213> *Saccharomyces cerevisiae*

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (1)...(1272)

&lt;223&gt; nucleotide sequence encoding Flip recombinase

&lt;400&gt; 60

```

atg cca caa ttt ggt ata tta tgt aaa aca cca cct aag gtg ctt gtt
Met Pro Gln Phe Gly Ile Leu Cys Lys Thr Pro Pro Lys Val Leu Val
1      5      10
cgt cag ttt gtg gaa agg ttt gaa aga cct tca ggt gag aaa ata gca
Arg Gln Phe Val Glu Arg Phe Glu Arg Pro Ser Gly Glu Lys Ile Ala
      20      25      30

```

48

96

-31-

tta	tgt	gct	gct	gaa	cta	acc	tat	tta	tgt	tgg	atg	att	aca	cat	aac	144
Leu	Cys	Ala	Ala	Glu	Leu	Thr	Tyr	Leu	Cys	Trp	Met	Ile	Thr	His	Asn	
		35					40					45				
gga	aca	gca	atc	aag	aga	gcc	aca	ttc	atg	agc	tat	aat	act	atc	ata	192
Gly	Thr	Ala	Ile	Lys	Arg	Ala	Thr	Phe	Met	Ser	Tyr	Asn	Thr	Ile	Ile	
	50					55					60					
agc	aat	tcg	ctg	agt	ttc	gat	att	gtc	aat	aaa	tca	ctc	cag	ttt	aaa	240
Ser	Asn	Ser	Leu	Ser	Phe	Asp	Ile	Val	Asn	Lys	Ser	Leu	Gln	Phe	Lys	
65					70					75					80	
tac	aag	acg	caa	aaa	gca	aca	att	ctg	gaa	gcc	tca	tta	aag	aaa	ttg	288
Tyr	Lys	Thr	Gln	Lys	Ala	Thr	Ile	Leu	Glu	Ala	Ser	Leu	Lys	Lys	Leu	
				85					90					95		
att	cct	gct	tgg	gaa	ttt	aca	att	att	cct	tac	tat	gga	caa	aaa	cat	336
Ile	Pro	Ala	Trp	Glu	Phe	Thr	Ile	Ile	Pro	Tyr	Tyr	Gly	Gln	Lys	His	
			100					105					110			
caa	tct	gat	atc	act	gat	att	gta	agt	agt	ttg	caa	tta	cag	ttc	gaa	384
Gln	Ser	Asp	Ile	Thr	Asp	Ile	Val	Ser	Ser	Leu	Gln	Leu	Gln	Phe	Glu	
		115					120					125				
tca	tcg	gaa	gaa	gca	gat	aag	gga	aat	agc	cac	agt	aaa	aaa	atg	ctt	432
Ser	Ser	Glu	Glu	Ala	Asp	Lys	Gly	Asn	Ser	His	Ser	Lys	Lys	Met	Leu	
	130					135					140					
aaa	gca	ctt	cta	agt	gag	ggg	gaa	agc	atc	tgg	gag	atc	act	gag	aaa	480
Lys	Ala	Leu	Leu	Ser	Glu	Gly	Glu	Ser	Ile	Trp	Glu	Ile	Thr	Glu	Lys	
145					150					155					160	
ata	cta	aat	tcg	ttt	gag	tat	act	tcg	aga	ttt	aca	aaa	aca	aaa	act	528
Ile	Leu	Asn	Ser	Phe	Glu	Tyr	Thr	Ser	Arg	Phe	Thr	Lys	Thr	Lys	Thr	
				165					170					175		
tta	tac	caa	ttc	ctc	ttc	cta	gct	act	ttc	atc	aat	tgt	gga	aga	ttc	576
Leu	Tyr	Gln	Phe	Leu	Phe	Leu	Ala	Thr	Phe	Ile	Asn	Cys	Gly	Arg	Phe	
			180					185					190			
agc	gat	att	aag	aac	gtt	gat	ccg	aaa	tca	ttt	aaa	tta	gtc	caa	aat	624
Ser	Asp	Ile	Lys	Asn	Val	Asp	Pro	Lys	Ser	Phe	Lys	Leu	Val	Gln	Asn	
		195					200					205				
aag	tat	ctg	gga	gta	ata	atc	cag	tgt	tta	gtg	aca	gag	aca	aag	aca	672
Lys	Tyr	Leu	Gly	Val	Ile	Ile	Gln	Cys	Leu	Val	Thr	Glu	Thr	Lys	Thr	
	210					215					220					
agc	gtt	agt	agg	cac	ata	tac	ttc	ttt	agc	gca	agg	ggg	agg	atc	gat	720
Ser	Val	Ser	Arg	His	Ile	Tyr	Phe	Phe	Ser	Ala	Arg	Gly	Arg	Ile	Asp	
225					230					235					240	
cca	ctt	gta	tat	ttg	gat	gaa	ttt	ttg	agg	aat	tct	gaa	cca	gtc	cta	768
Pro	Leu	Val	Tyr	Leu	Asp	Glu	Phe	Leu	Arg	Asn	Ser	Glu	Pro	Val	Leu	
				245					250					255		
aaa	cga	gta	aat	agg	acc	ggc	aat	tct	tca	agc	aat	aaa	cag	gaa	tac	816
Lys	Arg	Val	Asn	Arg	Thr	Gly	Asn	Ser	Ser	Ser	Asn	Lys	Gln	Glu	Tyr	
			260					265					270			
caa	tta	tta	aaa	gat	aac	tta	gtc	aga	tcg	tac	aat	aaa	gct	ttg	aag	864
Gln	Leu	Leu	Lys	Asp	Asn	Leu	Val	Arg	Ser	Tyr	Asn	Lys	Ala	Leu	Lys	
			275				280					285				
aaa	aat	gcg	cct	tat	tca	atc	ttt	gct	ata	aaa	aat	ggc	cca	aaa	tct	912
Lys	Asn	Ala	Pro	Tyr	Ser	Ile	Phe	Ala	Ile	Lys	Asn	Gly	Pro	Lys	Ser	
	290					295					300					

-32-

cac att gga aga cat ttg atg acc tca ttt ctt tca atg aag ggc cta His Ile Gly Arg His Leu Met Thr Ser Phe Leu Ser Met Lys Gly Leu 305 310 315 320	960
acg gag ttg act aat gtt gtg gga aat tgg agc gat aag cgt gct tct Thr Glu Leu Thr Asn Val Val Gly Asn Trp Ser Asp Lys Arg Ala Ser 325 330 335	1008
gcc gtg gcc agg aca acg tat act cat cag ata aca gca ata cct gat Ala Val Ala Arg Thr Thr Tyr Thr His Gln Ile Thr Ala Ile Pro Asp 340 345 350	1056
cac tac ttc gca cta gtt tct cgg tac tat gca tat gat cca ata tca His Tyr Phe Ala Leu Val Ser Arg Tyr Tyr Ala Tyr Asp Pro Ile Ser 355 360 365	1104
aag gaa atg ata gca ttg aag gat gag act aat cca att gag gag tgg Lys Glu Met Ile Ala Leu Lys Asp Glu Thr Asn Pro Ile Glu Glu Trp 370 375 380	1152
cag cat ata gaa cag cta aag ggt agt gct gaa gga agc ata cga tac Gln His Ile Glu Gln Leu Lys Gly Ser Ala Glu Gly Ser Ile Arg Tyr 385 390 395 400	1200
ccc gca tgg aat ggg ata ata tca cag gag gta cta gac tac ctt tca Pro Ala Trp Asn Gly Ile Ile Ser Gln Glu Val Leu Asp Tyr Leu Ser 405 410 415	1248
tcc tac ata aat aga cgc ata taa Ser Tyr Ile Asn Arg Arg Ile *	1272
420	

&lt;210&gt; 61

&lt;211&gt; 422

&lt;212&gt; PRT

<213> *Saccharomyces cerevisiae*

&lt;400&gt; 61

Pro Gln Phe Gly Ile Leu Cys Lys Thr Pro Pro Lys Val Leu Val Arg 1 5 10 15
Gln Phe Val Glu Arg Phe Glu Arg Pro Ser Gly Glu Lys Ile Ala Leu 20 25 30
Cys Ala Ala Glu Leu Thr Tyr Leu Cys Trp Met Ile Thr His Asn Gly 35 40 45
Thr Ala Ile Lys Arg Ala Thr Phe Met Ser Tyr Asn Thr Ile Ile Ser 50 55 60
Asn Ser Leu Ser Phe Asp Ile Val Asn Lys Ser Leu Gln Phe Lys Tyr 65 70 75 80
Lys Thr Gln Lys Ala Thr Ile Leu Glu Ala Ser Leu Lys Lys Leu Ile 85 90 95
Pro Ala Trp Glu Phe Thr Ile Ile Pro Tyr Tyr Gly Gln Lys His Gln 100 105 110
Ser Asp Ile Thr Asp Ile Val Ser Ser Leu Gln Leu Gln Phe Glu Ser 115 120 125
Ser Glu Glu Ala Asp Lys Gly Asn Ser His Ser Lys Lys Met Leu Lys 130 135 140
Ala Leu Leu Ser Glu Gly Glu Ser Ile Trp Glu Ile Thr Glu Lys Ile 145 150 155 160
Leu Asn Ser Phe Glu Tyr Thr Ser Arg Phe Thr Lys Thr Lys Thr Leu 165 170 175
Tyr Gln Phe Leu Phe Leu Ala Thr Phe Ile Asn Cys Gly Arg Phe Ser 180 185 190
Asp Ile Lys Asn Val Asp Pro Lys Ser Phe Lys Leu Val Gln Asn Lys 195 200 205
Tyr Leu Gly Val Ile Ile Gln Cys Leu Val Thr Glu Thr Lys Thr Ser

-33-

210	215	220
Val Ser Arg His Ile Tyr	Phe Phe Ser Ala Arg	Gly Arg Ile Asp Pro
225	230	235
Leu Val Tyr Leu Asp	Glu Phe Leu Arg Asn	Ser Glu Pro Val Leu Lys
245	250	255
Arg Val Asn Arg Thr	Gly Asn Ser Ser Ser	Asn Lys Gln Glu Tyr Gln
260	265	270
Leu Leu Lys Asp Asn	Leu Val Arg Ser Tyr	Asn Lys Ala Leu Lys Lys
275	280	285
Asn Ala Pro Tyr Ser	Ile Phe Ala Ile Lys	Asn Gly Pro Lys Ser His
290	295	300
Ile Gly Arg His Leu	Met Thr Ser Phe Leu	Ser Met Lys Gly Leu Thr
305	310	315
Glu Leu Thr Asn Val	Gly Asn Trp Ser Ser	Asp Lys Arg Ala Ser Ala
325	330	335
Val Ala Arg Thr Thr	Tyr Thr His Gln Ile	Thr Ala Ile Pro Asp His
340	345	350
Tyr Phe Ala Leu Val	Ser Arg Tyr Tyr Ala	Tyr Asp Pro Ile Ser Lys
355	360	365
Glu Met Ile Ala Leu	Lys Asp Glu Thr Asn	Pro Ile Glu Glu Trp Gln
370	375	380
His Ile Glu Gln Leu	Lys Gly Ser Ala Glu	Gly Ser Ile Arg Tyr Pro
385	390	395
Ala Trp Asn Gly Ile	Ile Ser Gln Glu Val	Leu Asp Tyr Leu Ser Ser
405	410	415
Tyr Ile Asn Arg Arg	Ile	
420		

<210> 62  
 <211> 48  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> IR2

<400> 62  
 gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttc

48

<210> 63  
 <211> 48  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> IR1

<400> 63  
 gaagttccta tactttctag agaataggaa cttcggaata ggaacttc

48

<210> 64  
 <211> 66  
 <212> DNA  
 <213> Bacteriophage mu

<220>  
 <221> CDS  
 <222> (1)...(66)  
 <223> nucleotide sequence encoding GIN recombinase

<400> 64  
 tca act ctg tat aaa aaa cac ccc gcg aaa cga gcg cat ata gaa aac  
 Ser Thr Leu Tyr Lys Lys His Pro Ala Lys Arg Ala His Ile Glu Asn  
 1 5 10 15

48

gac gat cga atc aat taa  
 Asp Asp Arg Ile Asn \*

66

-34-

20

<210> 65  
 <211> 21  
 <212> PRT  
 <213> bacteriophage mu

<400> 65  
 Ser Thr Leu Tyr Lys Lys His Pro Ala Lys Arg Ala His Ile Glu Asn  
 1 5 10 15  
 Asp Asp Arg Ile Asn  
 20

<210> 66  
 <211> 69  
 <212> DNA  
 <213> Bacteriophage mu

<220>  
 <221> CDS  
 <222> (1)...(69)  
 <223> nucleotide sequence encoding Gin recombinase

<400> 66  
 tat aaa aaa cat ccc gcg aaa cga acg cat ata gaa aac gac gat cga 48  
 Tyr Lys Lys His Pro Ala Lys Arg Thr His Ile Glu Asn Asp Asp Arg  
 1 5 10 15  
 atc aat caa atc gat cgg taa 69  
 Ile Asn Gln Ile Asp Arg \*  
 20

<210> 67  
 <211> 22  
 <212> PRT  
 <213> bacteriophage mu

<220>  
 <223> Gin recombinase of bacteriophage mu

<400> 67  
 Tyr Lys Lys His Pro Ala Lys Arg Thr His Ile Glu Asn Asp Asp Arg  
 1 5 10 15  
 Ile Asn Gln Ile Asp Arg  
 20

<210> 68  
 <211> 555  
 <212> DNA  
 <213> Escherichia coli

<220>  
 <221> CDS  
 <222> (1)...(555)  
 <223> nucleotide sequence encoding PIN recombinase

<400> 68  
 atg ctt att ggc tat gta cgc gta tca aca aat gac cag aac aca gat 48  
 Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp  
 1 5 10 15  
 cta caa cgt aat gcg ctg aac tgt gca gga tgc gag ctg att ttt gaa 96  
 Leu Gln Arg Asn Ala Leu Asn Cys Ala Gly Cys Glu Leu Ile Phe Glu  
 20 25 30



-35-

gac aag ata agc ggc aca aag tcc gaa agg ccg gga ctg aaa aaa ctg	144
Asp Lys Ile Ser Gly Thr Lys Ser Glu Arg Pro Gly Leu Lys Lys Leu	
35 40 45	
ctc agg aca tta tcg gca ggt gac act ctg gtt gtc tgg aag ctg gat	192
Leu Arg Thr Leu Ser Ala Gly Asp Thr Leu Val Val Trp Lys Leu Asp	
50 55 60	
cgg ctg ggg cgt agt atg cgg cat ctt gtc gtg ctg gtg gag gag ttg	240
Arg Leu Gly Arg Ser Met Arg His Leu Val Val Leu Val Glu Glu Leu	
65 70 75 80	
cgc gaa cga ggc atc aac ttt cgt agt ctg acg gat tca att gat acc	288
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr	
85 90 95	
agc aca cca atg gga cgc ttt ttc ttt cat gtg atg ggt gcc ctg gct	336
Ser Thr Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala	
100 105 110	
gaa atg gag cgt gaa ctg att gtt gaa cga aca aaa gct gga ctg gaa	384
Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Lys Ala Gly Leu Glu	
115 120 125	
act gct cgt gca cag gga cga att ggt gga cgt cgt ccc aaa ctt aca	432
Thr Ala Arg Ala Gln Gly Arg Ile Gly Gly Arg Arg Pro Lys Leu Thr	
130 135 140	
cca gaa caa tgg gca caa gct gga cga tta att gca gca gga act cct	480
Pro Glu Gln Trp Ala Ala Gly Arg Leu Ile Ala Ala Gly Thr Pro	
145 150 155 160	
cgc cag aag gtg gcg att atc tat gat gtt ggt gtg tca act ttg tat	528
Arg Gln Lys Val Ala Ile Ile Tyr Asp Val Gly Val Ser Thr Leu Tyr	
165 170 175	
aag agg ttt cct gca ggg gat aaa taa	555
Lys Arg Phe Pro Ala Gly Asp Lys *	
180	

&lt;210&gt; 69

&lt;211&gt; 184

&lt;212&gt; PRT

&lt;213&gt; Escherichia coli

&lt;400&gt; 69

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp	
1 5 10 15	
Leu Gln Arg Asn Ala Leu Asn Cys Ala Gly Cys Glu Leu Ile Phe Glu	
20 25 30	
Asp Lys Ile Ser Gly Thr Lys Ser Glu Arg Pro Gly Leu Lys Lys Leu	
35 40 45	
Leu Arg Thr Leu Ser Ala Gly Asp Thr Leu Val Val Trp Lys Leu Asp	
50 55 60	
Arg Leu Gly Arg Ser Met Arg His Leu Val Val Leu Val Glu Glu Leu	
65 70 75 80	
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr	
85 90 95	
Ser Thr Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala	
100 105 110	
Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Lys Ala Gly Leu Glu	
115 120 125	
Thr Ala Arg Ala Gln Gly Arg Ile Gly Gly Arg Arg Pro Lys Leu Thr	
130 135 140	
Pro Glu Gln Trp Ala Gln Ala Gly Arg Leu Ile Ala Ala Gly Thr Pro	
145 150 155 160	

-36-

Arg Gln Lys Val Ala Ile Ile Tyr Asp Val Gly Val Ser Thr Leu Tyr  
 165 170 175  
 Lys Arg Phe Pro Ala Gly Asp Lys  
 180

<210> 70  
 <211> 4778  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pcx plasmid

<400> 70  
 gtcgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60  
 gcccataatatt ggagttccgc gttacataaac ttacgggtaaa tggcccgccct ggctgaccgc 120  
 ccaacgacccc ccgccatttg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180  
 ggaacttttcca ttgacgtcaa tgggtgggact atttacggta aactgcccac ttggcagtag 240  
 atcaagtgtat tcatatgccca agtacgcccc ctattgacgt caatgacggg aaatggcccc 300  
 cctggcatta tgcccagtag atgaccttat gggactttcc tacttggcag tacatctacg 360  
 tattagtcac cgtattacc atgggtcgag gtgagcccc cgttctgctt cactctcccc 420  
 atctcccccc cctccccacc cccaattttt tatattatta ttttttaatt attttgtgca 480  
 gcgatggggg cggggggggg gggggcgccg gccaggcggg gcggggcggg gcgagggggc 540  
 gggcgggggc aggcggagag gtgcggcgcc agccaatcag agcggcgccg tccgaaagt 600  
 tccttttatg gcgagcgccg ggcggcgccg gccctataaa aagcgaagcg cgcggcgccc 660  
 gggagtcgct gcgttgccct cgcccccgtc ccgcctccgc gccgcctcgc gccgccccgc 720  
 ccgctctga ctgaccgct tactccaca ggtgagcggg cgggacggcc cttctctccc 780  
 gggctgtaat tagcgcttg ttaattgacg gctcgtttct tttctgtggc tgcgtgaaag 840  
 ccttaaaggc ctccgggagg gccctttgtg cggggggggg cggctcgggg ggtgctgctg 900  
 tgtgtgtgtg cgtggggagg gcccgctgcg gcccgcgctg cccggcggtg gtgagcgctg 960  
 cgggcgcggc gcggggcttt gtgcgctccg cgtgtgcgcg aggggagcgc gccggggggc 1020  
 ggtgccccgc ggtgcggggg ggctgcgagg ggaacaaagg ctgctgctgc ggtgtgtgctg 1080  
 tgggggggtg agcaggggtg gtggcgccgg cggtcgggct gtaaccccc cctgcacccc 1140  
 cctccccgag ttgctgagca cggccccggt tcgggtgcgg ggctccgtgc ggggctgggc 1200  
 gcggggctcg ccgtgccggg cgggggggtg cggcaggtgg ggggtgccgg cggggcgggg 1260  
 ccgcctcggg ccggggaggg ctccggggag ggcgcggcg gccccggagc gccggcggt 1320  
 gtcgagcgcc gccgagccgc agccattgoc ttttatggtat atcgtgcgag agggcgccag 1380  
 gacttccctt gtcccaaatc tggcggagcc gaaatctggg aggcgcggcc gcacccctc 1440  
 tagcggcgcc ggcgaagcg cgcgcgcgcg tccccttctc catctccagc ctccgggctg ccgcaggggg 1500  
 cgtgcgtcgc tcggggggga cggggcaggg cggggttcgg cttctggcgt gtgaccggcg 1620  
 acgtgctggt tctgctaacc catgttcatg ccttctctct tttctacag ctccgggca 1680  
 atgcctatca gaaggtggtg gctggtgtgg ccaatgccct ggctcacaaa taccactgag 1740  
 atctttttcc cctgtccaaa aattatgggg acatcatgaa gcccttgag catctgact 1800  
 ctggctaata aaggaaattt attttcattg caatagtgtg ttggaatttt ttgtgtctct 1860  
 cactcggaag gacatatggg agggcaaatc atttaaaaca tcagaatgag tatttggttt 1920  
 agagtttggc aacatatgcc atatgctggc tgccatgaac tccatagaaa agccttgact 1980  
 catcagtaga tgaggttaga ttttttttat attttgtttt gtgttatatt tttctttaac atccctaaaa 2040  
 ttttccttac atgttttact agccagattt ttctctctct cctgactact cccagtcata 2100  
 gctgtccctc ttctcttatg aagatccctc gacctgcagc ccaagcttgg cgtaatcatg 2160  
 gtcatagctg tttcctgtgt gaaattgtta tccgctcaca attccacaca acatacagagc 2220  
 cggaagcata aagtgtaaag cctgggggtg ctaatgagtg agctaactca cattaattgc 2280  
 gttgcgtcga ctgcccgtt tccagtcggg aaacctgtcg tgccagcgga tccgcattctc 2340  
 aattagtcag caaccatagt cccgccccta actccgcccc tcccgcccc aactccgccc 2400  
 agttccgccc attctccgcc ccattggctga ctaatttttt ttatttatgc agaggccgag 2460  
 gccgcctcgg cctctgagct attccagaag tagtgaggag gcttttttgg aggcctaggc 2520  
 ttttgcaaaa agctaacttg tttattgcag cttataatgg ttacaaataa agcaatagca 2580  
 tcacaaattt cacaaataaa gcattttttt cactgcattc tagttgtggt ttgtccaaac 2640  
 tcatcaatgt atcttatcat gtctggatcc gctgcattaa tgaatcgggg aacgcgcggg 2700  
 gagaggcggt ttgcgtattg ggcgctcttc cgtctcctcg ctactgact cgctgcgtc 2760  
 ggtcgttcgg tgcggcgag ggaatcagc tctactcaaa gcgtaatac ggttatccac 2820  
 agaatacagg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggcagggaa 2880  
 ccgtaaaaaa gccgcgttgc tggcgttttt ccataggctc cgccccctg acgagcatca 2940  
 caaaaatcga cgctcaagtc agaggtggcg aaaccgcaga ggactataaa gataccaggc 3000  
 gtttccccct ggaagctccc tcgtgcgtc tcctgttccg accctgccgc ttaccggata 3060  
 3120  
 3180

-37-

cctgtccgcc	ttttccctt	cggaagcgt	ggcgctttct	caatgctcac	gctgtaggta	3240
tctcagttcg	gtgtaggctg	ttcgctccaa	gctgggctgt	gtgcacgaac	cccccgttca	3300
gcccagaccg	tgcgccttat	ccggtacta	tgcgtcttgag	tccaacccgg	taagacacga	3360
cttatcgcca	ctggcagcag	ccactggtaa	caggattagc	agagcgagggt	atgtaggcgg	3420
tgctacagag	ttcttgaagt	ggtggcctaa	ctacggctac	actagaagga	cagtatttgg	3480
tatctgcgct	ctgctgaagc	cagttacctt	cggaaaaaga	gttggttagct	cttgatccgg	3540
caaaaaaacc	accgctggta	gcggtgggtt	ttttgtttgc	aagcagcaga	ttacgcgcag	3600
aaaaaaagga	tctcaagaag	atcctttgat	cttttctacg	gggtctgacg	ctcagtgga	3660
cgaaaactca	cgttaaggga	ttttgggtcat	gagattatca	aaaaggatct	tcacctagat	3720
ccttttaaat	taaaaatgaa	gttttaaatc	aatctaaagt	atatatgagt	aaacttggtc	3780
tgacagttac	caatgcttaa	tcagtgaggc	acctatctca	gcgatctgtc	tatttcgttc	3840
atccatagtt	gcctgactcc	ccgtcgtgta	gataactacg	atacgggagg	gcttaccatc	3900
tggccccagt	gctgcaatga	taccgcgaga	cccacgctca	ccggctccag	atztatcagc	3960
aataaaccag	ccagccggaa	gggcccagcg	cagaagggtt	cctgcaactt	tatccgcctc	4020
catccagttc	attaattggt	gccgggaagc	tagagtaagt	agttcgccag	ttaatatgtt	4080
gcgcaacggt	gttgccattg	ctacaggcat	cgtgggtgta	cgctcgctcg	ttggtatggc	4140
ttcattcagc	tccggttccc	aacgatcaag	gcgagttaca	tgatcccca	tgtgtgcaa	4200
aaaagcgggt	agctccttcg	gtcctccgat	cgttgctaga	agtaagttgg	ccgcagtggt	4260
atcactcatg	gttatggcag	cactgcataa	ttctcttact	gtcatgccat	ccgtaagatg	4320
cttttctgtg	actgggtgag	actcaaccac	gtcattctga	gaatagtgtg	tgcggcgacc	4380
gagttgctct	tgcctggcgt	caatacggga	taataccgcg	ccacatagca	gaactttaaa	4440
agtgtctcat	attggaaaac	gttcttcggg	gcgaaaactc	tcaaggatct	taccgctggt	4500
gagatccagt	tcgatgtaac	ccactcgtgc	acccaactga	tcttcagcat	cttttacttt	4560
caccagcgtt	tctgggtgag	caaaaacagg	aaggcaaaat	gccgcaaaaa	aggggaataag	4620
ggcgacacgg	aaatggtgaa	tactcatact	cttccttttt	caatattatt	gaagcattta	4680
tcaggggttat	tgtctcatga	gcggatacat	atttgaatgt	atthagaaaa	ataaacaat	4740
aggggttccg	cgcacatttc	ccgaaaaagt	gccacctg			4778

&lt;210&gt; 71

&lt;211&gt; 5510

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pCXeGFP plasmid

&lt;400&gt; 71

gtcgacattg	attattgact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccataat	ggagttccgc	gttacataac	ttacggtaaa	tggcccgccct	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtag	240
atcaagtgtg	tcatatgcc	agtacgccc	ctattgacgt	caatgacggg	aaatggccc	300
cctggcatta	tgcccagtag	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattatgcat	cgctattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctcccc	420
atctccccc	ctccccacc	cccaattttg	tatttattta	ttttttaatt	attttgtgca	480
gcgatggggg	cggggggggg	gggggcgcgc	gccaggcggg	gcggggcggg	gcgagggggc	540
gggcccggcg	agggcgagag	gtgcggcggc	agccaatcag	agcggcgcgc	tccgaaagt	600
tccttttatg	gcgaggcggc	ggcggcgggc	gcccataaaa	aagcgaagcg	cgcggcgggc	660
gggagtcgct	gcgttgccct	cgccccgtgc	cccgtccgc	gcgcctcgc	gcgcgccgc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtagcggg	cgggacggcc	cttctcctcc	780
gggctgtaat	tagcgtttgg	tttaattgac	gctcgtttct	tttctgtggc	tgctgaaag	840
ccttaaaggg	ctccgggagg	gcccttttgt	cgggggggag	cggtcggggg	ggtgcgtgcg	900
tgtgtgtgtg	cgtggggagc	gccgcgtgcg	gcccgcgctg	ccggcgggct	gtgagcgctg	960
cgggcgcggc	gcggggcctt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggccgggggc	1020
gggtgccccg	gggtgccccg	ggctgcgagg	ggaacaaagg	ctgcgtgcgg	gggtgtgtgc	1080
tgggggggtg	agcagggggg	gtgggcgcgg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cggcccgcct	tcgggtgcgg	ggctccgtgc	ggggcgtggc	1200
gcggggcctc	ccgtgccggg	cgggggggtg	cggcaggtgg	gggtgccggg	cggggcgggg	1260
cgcgctcggg	ccggggaggg	ctcgggggag	gggcgcggcg	gcccgggagc	gcccggcggt	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	tttttggtg	atcgtgcgag	agggcgcagg	1380
gacttccttt	gtcccaaatc	tggcggagcc	gaaatctggg	aggcgcgcgc	gcacccctc	1440
tagcgggcgc	gggcgaagcg	gtgcggcgcc	ggcaggaagg	aaatgggcgg	ggagggcctt	1500
cgtgcgtcgc	cgcgcgcgcg	tccccttctc	catctccagg	ctcggggctg	ccgagggggg	1560
acggctgcct	tccgggggga	cggggcaggg	cgggggttcg	cttctggcgt	gtgacggcg	1620
gctctagagc	ctctgctaac	catgttcatt	ccttcttctt	tttcttacag	ctcctgggca	1680
acgtgctggg	tgtgtgctg	tctcatcatt	ttggcaaga	attcgccacc	atgggtgagc	1740
agggcgaggga	gctgttcacc	gggggtggtgc	ccatcctggt	cgagctggac	ggcgacgtaa	1800

-38-

acggccacaa	ggtcagcgtg	tcgggcgagg	gcgagggcga	tgccacctac	ggcaagctga	1860
ccctgaagtt	catctgcacc	acgggcaagc	tgcccggtgcc	ctggcccacc	ctcgtgacca	1920
ccctgacct	cggtgcag	tgcttcagcc	gctaccccg	ccacatgaag	cagcacgact	1980
tcttcaagtc	cgccatgcc	gaaggctacg	tccaggagcg	caccatcttc	ttcaaggacg	2040
acggcaacta	caagaccgcg	gcccaggtga	agttcgaggg	cgacaccctg	gtgaaccgca	2100
tcgagctgaa	gggcatcgac	ttcaaggagg	acggcaacat	cctggggcac	aagctggagt	2160
acaactacaa	cagccacaac	gtctatatca	tggccgacaa	gcagaagaac	ggcatcaagg	2220
tgaacttcaa	gatccgccc	aacatcgagg	acggcagcgt	gcagctcgcc	gaccactacc	2280
agcagaacac	ccccatcgcc	gacggccccg	tgctgtgtgc	cgacaaccac	tacctgagca	2340
cccagtcg	cctgagcaaa	gaccccaacg	agaagcgcg	tcacatgggt	ctgctggagt	2400
tcgtgaccgc	cgccgggatc	actctcgcca	tggacgagct	gtacaagtaa	gaattcactc	2460
ctcaggtgca	ggctgcctat	cagaagggtg	tggctgggtg	ggccaatgcc	ctggctcaca	2520
aataccactg	agatcttttt	ccctctgcca	aaaattatgg	ggacatcatg	aagccccttg	2580
agcatctgac	ttctggctaa	taaaggaaat	ttattttcat	tgcaatagtg	tggttgaatt	2640
ttttgtgtct	ctcactcgga	aggacatatg	ggagggcaca	tcatttaaaa	catcagaatg	2700
agtatttggg	ttagagtttg	gcaacatatg	ccatatgctg	gctgccatga	acaaagggtg	2760
ctataaagag	gtcatcagta	tatgaaacag	ccccctgctg	tccattcctt	attccataga	2820
aaagccttga	cttgaggtta	gatttttttt	atattttggt	ttgtgttatt	tttttcttta	2880
acatccctaa	aattttcctt	acatgtttta	ctagccagat	ttttcctcct	ctcctgacta	2940
ctcccagtc	tagctgtccc	totttcttta	tgaagatccc	tcgacctgca	gccccagctt	3000
ggcgtaatca	tggtcatagc	tgtttctgtg	gtgaaattgt	tatccgctca	caattccaca	3060
caacatacga	ggcggaagca	taaagtgtaa	agcctggggg	gcctaataag	tgagctaact	3120
cacattaatg	ggcttgcgct	cactgcccgc	tttccagctg	ggaaacctgt	cgtgcccagc	3180
gatccgcctc	tcaattagtc	agcaaccata	gtcccgcctc	taactccgcc	catcccgcct	3240
ctaaactccg	ccagttccgc	ccattctccg	ccccatggct	gactaatttt	ttttatttat	3300
gcagaggccg	aggccgcctc	ggcctctgag	ctattccaga	agtagtgagg	aggctttttt	3360
ggaggccctag	gcttttgcaa	aaagctaaat	tggtttatga	agcttataat	ggttacaaat	3420
aaagcaaatg	catcacaat	ttocaaaaat	aagcattttt	ttcactgcat	tctagttgtg	3480
gtttgtccaa	actcatcaat	gtatcttata	atgtctggat	ccgctgcatt	aatgaatcgg	3540
ccaacgcgcg	gggagaggcg	gtttgcgtat	tgggcgctct	tccgcttctt	cgtcactga	3600
ctcgtgcgcg	tcggtcgctc	ggctgcggcg	agcgggtatc	gctcactcaa	aggcggtaat	3660
acggttatcc	acagaatcag	gggataacgc	aggaaagaac	atgtgagcaa	aaggccagca	3720
aaaggccagg	aaccgtaaaa	aggccgcggt	gctggcggtt	ttccataggc	tccgcccccc	3780
tgacgagcat	cacaaaaatc	gacgtcaag	tcagaggtgg	cgaaaccgca	caggactata	3840
aagataccag	gcgtttcccc	ctggaagctc	cctcgtgcgc	tctcctgttc	cgaccctgcc	3900
gcttacccga	tacctgtccg	cccttctccc	ttcggaagc	gtggcgcttt	ctcaatgctc	3960
acgctgtagg	tactcagtt	cggtgtagg	cggtcgctcc	aagctggggt	gtgtgcacga	4020
accccccggt	cagcccgacc	gctgcgcctt	atccggtaac	tatcgtcttg	agtccaaccc	4080
ggtaagacac	gacttatcgc	cactggcagc	agccactggt	aacaggatta	gcagagcgag	4140
gtatgttagc	gggtctacag	agttcttgaa	gtggtggcct	aactacggct	acactagaag	4200
gacagtattt	gggtatctgc	ctctgctgaa	gccagttacc	ttcggaaaaa	gagttggtag	4260
ctcttgatcc	ggcaaacaaa	ccaccgctgg	tagcggtggg	ttttttgttt	gcaagcagca	4320
gattacgcgc	agaaaaaaag	gatctcaaga	atcctctttg	atcttttcta	cggggtctga	4380
cgctcagttg	aacgaaaact	cacgttaagg	gattttgggt	atgagattat	caaaaaggat	4440
cttcacctag	atccttttaa	attaaaaatg	aagttttaaa	tcaatctaaa	gtatatatga	4500
gtaaaacttg	tctgacagtt	accaatgctt	aatcagtgag	gcacctatct	cagcgatctg	4560
tctatttcgt	tcatccatag	ttgcctgact	ccccgctcgt	tagataacta	cgatacggga	4620
gggcttacca	tctggcccc	gtgctgcaat	gataccgcga	gacccacgct	caccggctcc	4680
agattttatc	gcaataaacc	agccagccgg	aaggccgag	cgcagaagtg	gtcctgcaac	4740
tttatccgcc	tccatccagt	ctattaattg	ttgcgggaa	gctagagtaa	gtagtccgcc	4800
agttaatagt	ttgcgcaacg	ttgttgccat	tgctacaggc	atcgtgggtg	cacgctcgct	4860
gtttgggtatg	gcttcattca	gctccgggtc	ccaacgatca	aggcgagtta	catgatcccc	4920
catgtttgtg	aaaaaagcgg	ttagctcctt	cggtcctccg	atcgttgtca	gaagtaagtt	4980
ggccgcagtg	ttatcactca	tggttatggc	agcactgcat	aattctctta	ctgtcatgce	5040
atccgtaaga	tgcttttctg	tgactgggtg	gtactcaacc	aagtcattct	gagaatagtg	5100
tatgcccgcga	ccgagttgct	cttgcccggc	gtcaatacgg	gataatacgg	cgccacatag	5160
cagaacttta	aaagtgtcca	tcattggaaa	acgtttcttc	gggcgaaaa	tctcaaggat	5220
cttaccgctg	ttgagatcca	gttcgatgta	accactcgt	gcaccaact	gatcttcagc	5280
atcttttact	ttcaccagcg	tttctgggtg	agcaaaaaa	ggaaggcaaa	atgccgcaaa	5340
aaagggtaata	agggcgacac	ggaaatgttg	aataactata	ctcttccttt	ttcaatatta	5400
ttgaagcatt	attcagggtt	attgtctcat	gagcggtatc	atatttgaat	gtatttagaa	5460
aaataaacaa	ataggggttc	cgcgcacatt	tccccgaaaa	gtgccacctg		5510

&lt;210&gt; 72

&lt;211&gt; 282

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

-39-

<220>  
<223> attp

<400> 72  
ccttgcgcta atgctctgtt acaggtcact aataccatct aagtagttga ttcatagtga 60  
ctgcatatgt tgtgtttttac agtattatgt agtctgtttt ttatgcaaaa tctaatttaa 120  
tatattgata tttatatcat tttacgtttc tcgttcagct tttttatact aagttggcat 180  
tataaaaaag cattgcttat caatttggtg caacgaacag gtcactatca gtcaaaataa 240  
aatcattatt tgatttcaat tttgtccac tccctgcctc tg 282

<210> 73  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 73  
ggccccgtaa tgcagaagaa 20

<210> 74  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 74  
ggtttaaagt gcgctcctcc aagaacgtca tc 32

<210> 75  
<211> 40  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 75  
agatctagag ccgcccgtac aggaacaggt ggtggcggcc 40

<210> 76  
<211> 37  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer 5PacSV40

<400> 76  
ctgttaatta actgtggaat gtgtgtcagt taggggtg 37

<210> 77  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer Antisense Zeo

<400> 77  
tgaacagggt cacgtcgtcc 20

<210> 78  
<211> 24

-40-

<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Primer 5' HETS  
  
<400> 78  
gggccgaaac gatctcaacc tatt 24  
  
<210> 79  
<211> 19  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Primer 3' HETS  
  
<400> 79  
cgcagcggcc ctctactc 19  
  
<210> 80  
<211> 29  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Primer 5BSD  
  
<400> 80  
accatgaaaa catttaacat ttctcaaca 29  
  
<210> 81  
<211> 29  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Primer SV40polyA  
  
<400> 81  
tttatttgtg aaatttgtga tgctattgc 29  
  
<210> 82  
<211> 25  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Primer 3BSP  
  
<400> 82  
ttaatttcgg gtatatttga gtgga 25  
  
<210> 83  
<211> 32  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Primer EPO5XBA  
  
<400> 83  
tatctagaat ggggggtgcac gaatgtcctg cc 32  
  
<210> 84  
<211> 32  
<212> DNA

-41-

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer EPO3SBI

&lt;400&gt; 84

tacgtacgtc atctgtcccc tgtcctgcag gc

32

&lt;210&gt; 85

&lt;211&gt; 27

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer GENEPO3BSI

&lt;400&gt; 85

cgtacgtcat ctgtcccctg tccctgca

27

&lt;210&gt; 86

&lt;211&gt; 28

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer GENEPO5XBA

&lt;400&gt; 86

tctagaatgg ggggtgcacgg tgagtact

28

&lt;210&gt; 87

&lt;211&gt; 4862

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pD2eGFP-1N plasmid from Clontech

&lt;400&gt; 87

tagttattaa	tagtaatcaa	ttacgggggtc	attagttcat	agcccatata	tggagttccg	60
cgttacataa	cttacggtaa	atggcccgcgc	tggctgaccg	cccaacgacc	cccggccatt	120
gacgtcaata	atgacgtatg	ttcccacatg	aacgccaata	gggactttcc	attgacgtca	180
atgggtggag	tatttacggg	aaactgcccc	cttggcagta	catcaaggtg	atcatatgcc	240
aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gocctggcatt	atgccagta	300
catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac	360
catgggtgatg	cggttttggc	agtacatcaa	tgggcgtgga	tagcgggttg	actcacgggg	420
atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	480
ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	gtaggcgtgt	540
acggtgggag	gtctatatata	gcagagctgg	tttagtgaac	cgtcagatcc	gctagcgcta	600
ccggaactcag	atctcgagct	caagcttcga	attctgcagt	cgacgggtacc	gcgggcccg	660
gatccaccgg	tcgccaccat	ggtagcaag	ggcgaggagc	tgttcaccgg	ggtaggtgcc	720
atcctggtcg	agctggacgg	cgacgtaaac	ggccacaagt	tcagcgtgtc	cggcgagggc	780
gagggcgatg	ccacctacgg	caagctgacc	ctgaagttca	tctgcaccac	cggcaagctg	840
cccgtgccct	ggcccaccct	cgtgaccacc	ctgacctacg	gcgtgcagtg	cttcagccgc	900
taccocgacc	acatgaagca	gcacgacttc	ttcaagtccg	ccatgcccca	aggctacgtc	960
caggagcgca	ccatcttctt	caaggacgac	ggcaactaca	agaccgcgcg	cgaggtgaag	1020
ttcgagggcg	acaccctggg	gaaccgcac	gagctgaagg	gcacgcactt	caaggaggac	1080
ggcaacatcc	tggggcacaa	gctggagtag	aactacaaca	gccacaacgt	ctatatcatg	1140
gccgacaagc	agaagaacgg	catcaagggtg	aacttcaaga	tccgccacaa	catcgaggac	1200
ggcagcgtgc	agctcgccga	ccactaccag	cagaacaccc	ccatcgccga	cggccccctg	1260
ctgtgcocg	acaaccacta	cctgagcacc	cagtcgcgcc	tgagcaaaga	ccccaacgag	1320
aagcgcgatc	acatggtcct	gctggagttc	gtgaccgcgc	ccgggatcac	tctcggcatg	1380
gacgagctgt	acaagaagct	tagccatggc	ttcccgcggg	aggtggagga	gcaggatgat	1440
ggcacgctgc	ccatgtcttg	tgcccaggag	agcgggatgg	accgtcacc	tgcagcctgt	1500
gcttctgcta	ggatcaatgt	gtagatgcgc	ggcgcgcact	ctagatcata	atcagccata	1560
ccacatttgt	agagggttta	cttgctttta	aaaacctccc	acacctcccc	ctgaacctga	1620
aacataaaat	gaatgcaatt	gttgttggtta	acttgtttat	tgcagcttat	aatggttaca	1680

-42-

```
aataaagcaa tagcatcaca aatttcacaa ataaagcatt tttttcactg cattctagtt 1740
gtggtttgtc caaactcatc aatgtatctt aaggcgtaaa ttgtaagcgt taatatatttg 1800
ttaaaattcg cgttaaattt ttgttaaate agctcatttt ttaaccaata ggccgaaatc 1860
ggcaaaatcc cttataaatc aaaagaatag accgagatag ggttgagtgt tgttccagtt 1920
tggaacaaga gtccactatt aaagaacgtg gactccaacg tcaaagggcg aaaaaccgtc 1980
tatcagggcg atggccact acgtgaacca tcaccctaata caagtttttt ggggtcgagg 2040
tgccgtaaaag cactaaatcg gaaccctaaa gggagccccc aagtagagc ttgacgggga 2100
aagccggcga acgtggcgag aaaggaaggg aagaaagcga aaggagcggg cgctagggcg 2160
ctggcaagtg tagcgggtcac gctgcgcgta accaccacac ccgccgcgct taatgcgccg 2220
ctacagggcg cgtcaggtgg cacttttcgg ggaatgtgc gcggaacccc tatttgttta 2280
tttttctaaa tacattcaaa tatgtatccg ctcatgagac aataaccctg ataaatgctt 2340
caataatatt gaaaaaggaa gagtcttgag gcggaaagaa ccagctgtgg aatgtgtgtc 2400
agttaggggtg tggaaagtcc ccagggtccc cagcaggcag aagtagcaa agcatgcac 2460
tcaattagtc agcaaccagg tgtggaagt cccaggctc cccagcagg agaagtatgc 2520
aaagcatgca tctcaattag tcagcaacca tagtcccgcc cctaactccg cccatcccgc 2580
ccctaactcc gccaggttcc gagggccgcc ctgactaatt ttttttattt 2640
atgcagagggc caggccctcg acgtatcca gaagtagtga ggaggtttt 2700
ttggagccct aggccttttc aaagatcgat gatgaggatc gtttcgcatg 2760
attgaacaag atggattgca cgcaggttct cgggcccgtt ggttgagag gctattcgcc 2820
tatgactggg cacaacagac aatcgggtgc tctgatgcc cgtgttccg gctgtcaggc 2880
caggggcgcc cggttctttt tgtcaagacc gacctgtccg gtgccctgaa tgaactgcaa 2940
gacgaggcag cgcggtatc gtggctggcc acgacgggcg ttccttgccg agctgtgtc 3000
gacgttgtca ctgaaagcggg aagggactgg ctgctattgg gcgaagtgcc ggggcaggat 3060
ctcctgtcat ctacacttgc tctgcccag aaagtatcca tcatggctga tgcaatgcgg 3120
cggctgcata cgcttgatcc ggctacctgc ccattcgacc accaagcgaa acatcgcatc 3180
gagcgagcac gtactcggat ggaagccggg cttgtcgatc aggatgatc ggacgaagag 3240
catcaggggc tcgcgccagc cgaactgttc aggcagcatc aggcagcat gcccgacggc 3300
gaggatctcg tcgtgaccca tggcgatgcc tgcttgccga atatcatggt ggaaaatggc 3360
gcgttttctg gattcatcga ctgtggccgg ctgggtgtgg cggaccgcta tcaggacata 3420
gcgttggtta cccgtgatat tgctgaagag tctatgaaag aatgggctga ccgcttctc 3480
gtgctttacg gtatcgccgc tcccgattcg cagcgcatcg ccttctatcg ccttcttgac 3540
gagttcttct gagcgggact ctgggggttc aaatgaccga ccaagcgacg cccaacctgc 3600
catcacgaga ttctgattcc accgcgcctc gttgggcttc ggaatcggtt 3660
tccgggacgc cggctggatg atcctccagc gcggggatct catgctggag ttcttcgccc 3720
accctagggg gaggtctaat gaaacacgga agggagacaa accggaagga acccgcgcta 3780
tgacggcaat aaaaagacag aataaaacgc acggtgttgg gtcgtttgtt cataaacgcg 3840
gggttcgggtc ccagggtcgg cactctgtcg atacccacc gagaccccat tggggccaat 3900
acgcccgcgt ttcttccttt tccccacccc acccccgaag ttcgggtgaa ggcccagggc 3960
tcgcagccaa cgtcggggcg gcaggccctc cctggcgccg aggttactca tatatacttt 4020
agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata 4080
atatcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag 4140
aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa 4200
caaaaaaacc accgctacca gcggtggttt ttttgccgga tcaagagcta ccaactcttt 4260
ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc 4320
cgtagttagg ccaccacttc aagaactctg tagcacgcc tacataacct gctctgctaa 4380
tctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg ttggactcaa 4440
gacgatagtt accggataag gcgcagcggg tgagatacct ggggggttcg tgcacacagc 4500
ccagcttgga gcgaacgacc tacaccgaac acagcgtgag ctatgagaaa 4560
gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc ggttaagcggc agggctcgaa 4620
caggagagcg caccagggag cttccagggg gaaacgcctg gtatctttat agtccgtgcg 4680
ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc 4740
tatggaaaaa cgccagcaac gcggcctttt tacggttcct ggccttttgc tggccttttg 4800
ctcacatggt ctttcctgcg ttatcccctg attctgtgga taaccgtatt accgccatgc 4860
at
```

&lt;210&gt; 88

&lt;211&gt; 5192

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pIRESpuro2 plasmid from Clontech

&lt;400&gt; 88

```
gacggatcgg gagatctccc gatccccctat ggtcgactct cagtacaate tgctctgatg 60
ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gactagtgcg 120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180
```



-43-

ttaggggttag	gcgttttgcg	ctgcttcgcg	atgtacgggc	cagatatacg	cgttgacatt	240
gattattgac	tagttattaa	tagtaatcaa	ttacgggggc	attagttcat	agcccatata	300
tggagttccg	cgttacataa	cttacggtaa	atggcccgcg	tggctgaccg	cccaacgacc	360
cccgcccatt	gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	420
attgacgtca	atgggtggac	tatttacggg	aaactgcccc	cttggcagta	catcaagtgt	480
atcatatgcc	aagtacggcc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	540
atgcccagta	catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	600
tgcgtattac	catgggtgat	cggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	660
actcacgggg	atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	720
aaaatcaacg	ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	780
gtaggcgtgt	acgggtgggag	gtctatataa	gcagagctct	ctggctaact	agagaaccca	840
ctgcttactg	gcttatcgaa	attaatacga	ctcactatag	ggagacccaa	gcttgggtacc	900
gagctcggat	cgatatctgc	ggcctagcta	gcgcttaagg	cctgttaacc	ggtcgtacgt	960
ctcccgattc	gaattcggat	ccgcggccgc	atagataaact	gatccagtgt	gctggaatta	1020
attcgctgtc	tgcgagggcc	agctgttggg	gtgagtactc	cctctcaaaa	gcgggcatga	1080
cttctgcgct	aagattgtca	gtttccaaaa	acgaggagga	tttgatattc	acctggcccg	1140
cggtgatgcc	tttgaggggtg	gccgcgtcca	tctggtcaga	aaagacaatc	tttttgttgt	1200
caagcttgag	cttgaggcagg	cttgagatct	ggccatacac	ttgagtgaca	atgacatcca	1260
ctttgccttt	ctctccacag	gtgtccactc	ccaggtccaa	ctgcaggtcg	agcatgcatc	1320
tagggcggcc	aattccgccc	ctctccctcc	ccccccccta	acgttactgg	ccgaagccgc	1380
ttggaataag	gcgggtgtgc	gtttgtctat	atgtgatttt	ccaccatatt	gccgtctttt	1440
ggcaatgtga	gggcccggaa	acctggccct	gtcttcttga	cgagcattcc	taggggtctt	1500
tccctctcgc	ccaaaggaat	gcaaggtctg	ttgaatgtcg	tgaaggaaagc	agttcctctg	1560
gaagcttctt	gaagacaaac	aacgtctgta	gcgacccttt	gcaggcagcg	gaacccccca	1620
cctggcgaca	ggtgcctctg	cggccaaaag	ccacgtgtat	aagatacacc	tgcaaaggcg	1680
gcacaacccc	agtgccacgt	tgtgagttgg	atagttgtgg	aaagagtcaa	atgggtctcc	1740
tcaagcgtat	tcaacaaggg	gctgaaggat	gccagaagg	tacccatttg	tatgggatct	1800
gatctggggc	ctcgggtgcac	atgctttaca	tgtgtttagt	cgaggttaaa	aaaacgtcta	1860
ggccccccga	accacgggga	cgtgggttttc	ctttgaaaaa	cacgatgata	agcttggcac	1920
aacccacagc	gagacgacct	tccatgacct	agtacaagcc	cacggtgcgc	ctcggcacc	1980
gcgacgacgt	cccccgggcc	gtacgcaccc	tgcgcccgcc	gttcgcccgc	taccccgcca	2040
cgcgccacac	cgtcgacccc	gaccgcccaca	tcgagcgggt	caccgagctg	caagaactct	2100
tccctcacgcg	cgtcgggctc	gacatcggca	aggtgtgggt	cgccgacgac	ggcgcccgcg	2160
tggcggtctg	gaccacgcgc	gagagcgtcg	aagcgggggg	ggtgttcgcc	gagatcggcc	2220
cgcgcatggc	cgagttgagc	ggttcccggc	tggcccgcca	gcaacagatg	gaaggcctcc	2280
tggcgccgca	ccggcccga	gagcccgctg	ggttctctgg	caccgtcgcc	gtctcgcccg	2340
accaccaggg	caagggtctg	ggcagcgccg	tctgtctccc	cggagtggag	gcggccgagc	2400
gcgcgggggt	gcccgccttc	ctggagacct	ccgcgccccg	caacctcccc	ttctacgagc	2460
ggctcggtct	caccgtcacc	gccgacgtcg	agtgcccgaa	ggaccgcgcg	acctggtgca	2520
tgaccgcgca	cccggtgccc	tgacgcccgc	cccagacccc	gcagcgcccc	accgaaagga	2580
gcgcacgacc	ccatggctcc	gaccgaagcc	gaccggggcg	gccccgcccga	ccccgcaccc	2640
gcccccgagg	cccaccgact	ctagagctcg	ctgactcagc	tcgactgtgc	ttctagtgtc	2700
ccagccatct	gtttgttgcc	cctccccctg	gccttccctg	accctggaag	gtgccactcc	2760
cactgtcctt	tcctaataaa	atgaggaaat	tgcacgcgat	tgtctgagta	ggtgtcattc	2820
tattctgggg	ggtgggggtg	ggcaggacag	caagggggag	gattgggaag	acaatagcag	2880
gcatgctggg	gatgcgggtg	gctctatggc	ttctgaggcg	gaaagaacca	gctggggctc	2940
gagtgcatte	tagttgtggg	ttgtccaaac	tcatcaatgt	atcttatcat	gtctgtatac	3000
cgtcgacctc	tagctagagc	ttggcgtaat	catggtcata	gctgtttcct	gtgtgaaatt	3060
gttatccgct	cacaattcca	cacaacatac	gagccgggaag	cataaagtgt	aaagcctggg	3120
gtgcctaatt	agtgagctaa	ctcacattaa	ttgcgtttgcg	ctcactgccc	gctttccagt	3180
cgggaaacct	gtcgtgccag	ctgcattaat	gaatcgccca	acgcgcgggg	agaggcggtt	3240
tgcgtattgg	gcgctcttcc	gcttctctgc	tcactgactc	gctgcgctcg	gtcgttcggc	3300
tgccgcgagc	ggtatcagct	cactcaaaag	cggtaatacg	gttatccaca	gaatcagggg	3360
ataacgcagg	aaagaacatg	tgagcaaaa	gccagcaaaa	ggccaggaac	cgtaaaaagg	3420
ccgcgttgct	cataggctcc	gccccctga	cgagcatcac	cgagcatcac	aaaaatcgac	3480
gctcaagtca	gaggtggcga	gactataaag	ataccaggcg	ataccaggcg	tttccccctg	3540
gaagctccct	cgtgcgctct	ccctgccgct	taccggatac	taccggatac	ctgtccgctc	3600
ttctcccttc	gggaagcgtg	gcgctttctc	ctgtagggtat	ctgtagggtat	ctcagttcgg	3660
tgtaggtcgt	tcgctccaag	ctgggctgtg	tgcacgaacc	ccccgttcag	ccccgacctg	3720
tgcccttatc	cggtaaactat	cgtcttgagt	ccaacccggg	aagacacgac	ttatcgccac	3780
tggcgcagcg	cactggtaac	aggattagca	gagcgaggta	tgtaggcggt	gtcacagagt	3840
tcttgaagtg	gtggccctaac	tacggctaca	ctagaaggac	agtatttggg	atctgcgctc	3900
tgtgaaagcc	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	aaacaaacca	3960
ccgctggtag	cggtgttttt	tttgtttgca	agcagcagat	tacgcgcaga	aaaaaaggat	4020
ctcaagaaga	tccttttgatc	ttttctacgg	ggtctgacgc	tcagtggaaac	gaaaactcac	4080
gttaagggat	tttgggtcatg	agattatcaa	aaaggatctt	cacctagatc	cttttaaat	4140
aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttgggtc	gacagttacc	4200

-44-

```

aatgcttaat cagtgaggca cctatctcag cgatctgtct atttcgttca tccatagtgtg 4260
cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct ggccccagtg 4320
ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca ataaaccagc 4380
cagccggaag ggccgagcgc agaagtgggtc ctgcaacttt atccgcctcc atccagttcta 4440
ttaattgttg ccgggaagct agagtaagta gttcgccagt taatagtgtg cgcaacgttg 4500
ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggtagggct tcattcagct 4560
ccggttccca acgatcaagg cgagttacat gatcccccac gtgtgtgcaaa aaagcgggta 4620
gttccttcgg tcctccgatc gttgtcagaa gtaagtgggc cgcagtgtta tcaactcatgg 4680
ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc ttttctgtga 4740
ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg agttgctctt 4800
gcccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa gtgctcatca 4860
ttggaaaaac ttcttcgggg cgaaaactct caaggatctt accgctgttg agatccagtt 4920
cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc accagcgttt 4980
ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa gggaataaagg gcgacacgga 5040
aatgttgaat actcatactc ttcctttttc aatattattg aagcatttat cagggttatt 5100
gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata ggggttccgc 5160
gcacatttcc ccgaaaaagt ccacctgacg tc 5192

```

<210> 89  
 <211> 11182  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pAg1 Plasmid

```

<400> 89
catgccaaacc acagggttcc cctcgggatac aaagtacttt gatccaaccc ctccgctgct 60
atagtgcagt cggcttctga cgttcagtgac agccgtcttc tgaaaaacgc atgtcgcaca 120
agtcctaagt tacgcgcagc gctgccgcc tgcctttttc ctggcgtttt ttgtcgcgt 180
gttttagtcg cataaagtag aatacttgcc actagaaccg gagacattac gccatgaaca 240
agagcgccgc cgctggcctg ctgggctatg ccgcgctcag caccgacgac caggacttga 300
ccaaccaacg ggccgaactg cacgcggccg gctgcaccaa gctgttttcc gagaagatca 360
ccggcaccag ccggagctgg ccggagctgg ctagaccgcc tggcccgcag cctggcagag 420
acgttgtgac agtgaccagg gccggcgccg cctgcgtag cgtggcagag ccgtgggccc 480
ttgccgagcg catccaggag cgcattgggtg tgaccgtgtt cgccggcatt gccgagttcg 540
acaccaccac ccggcgccgc cgcacccgga gcggcgccga ggccgccaag gcccgaggcg 600
agcgttccct aatcatcgac accctcaccg cggcacagat cgcgcacgcc cgcgagctga 660
tgaagtttgg cccccgccct accctcaccg gtgaaagagg gcttggcgctg catcgctga 720
tcgaccagga aggcgcgacc cgcagcgagg aagtgcgccc caccgaggcc aggcggcgcg 780
ccctgtaccg cgcacttgag ttgaccgagg cccagccctt ggccggccgc gagaatgaac 840
gtgccttccg tgaggacgca aaccgcacca ggacggccag gacgaaccgt ttttcattac 900
gccaagagga acaagcatga tgatcgcgcc cgggtacgtg gatgccaagc tggcggcctg 960
cgaagagatc gaggcggaga aaatcctggc cgggttgtct ctaaaaaggt gatgtgtatt 1020
ctcaaccgtg cggctgcatg aagaaccgga gcgcccgtg atgcgatgag accagaaaag cgggtcagcg 1080
gcccggccagc ttggccgctg atgcggctcg gctgtactta tcccgggggc cgatgttctg 1140
tgagtaaaac agcttgctgc aagaaaccca ggcattggg cgcgattggg ggaagatcaa 1200
aatacgcaag gggaacgcat tctagccgc cgcgattggg cgcgactgaa ggccatcggc 1260
aagacgacca tcgcaaccca gggcagtgcc acgattgacc cggacttggc tgtgtccgcg 1320
ccgctaaccc ttgtcggcat cgacggagcg gctgatccg gtgacgcaa gcccttacga catatgggccc 1380
cggcgcgact tcgtagtgat cgacggagcg cgttaagcag cgcattgagg tcacggatgg aaggctacaa 1440
atcaaggcag ccgacttcgt gctgaaatt ggcgatcaaa ggcacgcgca tgggcccgtg ggttggccgag 1500
accgcccagc tggtagtgat gcccattctt cactgaatcag aaatcaaaac taatgaggta 1560
gcccgttggc cggccgcccg cacaaccgtt cgctgaaatt aagtgccgg ccgtccgagc gcacgcagca 1620
ccgagagctg aggcgctggc gagcaaaagc agcctggcag acacgcccag catgaagcgg gtcaactttc 1680
agtgccggcg ggaggtcac gctatctgaa tacatcgccg ggctaaagga ggaacggggc gttggccagg atcggcgctga 1740
ttaccgagct gaatttttag tggaaatccc caatggcact ggaacccccca agcccgagga acctgggtgga 1800
atgagtagat tggaaatccc caatggcact ggaacccccca agcccgagga acctgggtgga 1860
accgagcccg tggaaatccc caatggcact ggaacccccca agcccgagga acctgggtgga 1920
tgggttgtct cggtcgcaaa gggcgcgagc cggcccccgg cagccccggg 1980
cggtcgcaaa gggcgcgagc cggcccccgg cagccccggg cagccccggg 2040
gaagttgaag gggcgcgagc cggcccccgg cagccccggg cagccccggg 2100

```

tgaatcgtgg	caagcggccg	ctgatcgaat	cgcgaagaa	tcccggcaac	cgccggcagc	2460
cggtgcgcg	tcgattagga	agccgcacca	gggcgacgag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tcgcagcatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	gggtgatccgc	tacgagcttc	cagacgggca	2640
cgtlagaggtt	tcgcgagggc	cggccggcat	ggccagtgtg	tgggattacg	acctgggtact	2700
gatggcggtt	tcccatctaa	ccgaatccat	gaaccgatac	cggaaggga	agggagacaa	2760
gcccggccgc	gtgttcgctc	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaaa	cagaaagacg	acctggtaga	aaactgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggccgcctg	gtgacggtat	ccgaggggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	ggcgaaacc	gggcggccgg	agtagactcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggaagctgct	3060
gacggttcac	cccgattact	ttttgatcga	tcccggcatc	ggccgttttc	tctaccgcct	3120
ggcaccgcgc	gcccagggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gcccagagag	tcaagaagtt	ctgtttcacc	gtgcgcaagc	tgatcggttc	3240
aaatgacctg	ccggagtagc	atgttgaggga	ggaggcgggg	caggctggcc	cgatcctagt	3300
catgcgctac	cgaacacctga	tcgagggcga	agcatccgcc	ggttcctaata	gtacggagca	3360
gatgctaggg	caaatgtccc	tagcagggga	aaagggtcga	aaagggtcct	ttctgtgga	3420
tagcagctga	attgggaacc	caaagccgta	cattgggaac	cggaaccctg	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	catgtaagtg	actgatataa	aagagaaaaa	3540
aggcgatttt	tccgcctaaa	actctttaaa	acttattaaa	actcttaaaa	ccgcctggc	3600
ctgtgcataa	ctgtctggcc	agcgacagc	cgaagagctg	caaaaagcgc	ctaccttcg	3660
gtcgctgcgc	tccctacgcc	ccgcgccttc	gcgtcggcct	atcgcgccgc	ctggccgctc	3720
aaaaatggct	ggcctacggc	caggcaatct	accagggcgc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	cggcgccac	atcaaggcac	ctgcctcgc	gcgtttcggt	gtgacaggtg	3840
aaaacctctg	acacatgcag	ctcccggaga	cggtcacagc	ttgtctgtaa	gcggtgccc	3900
ggagcagaca	agcccgctcag	ggcgcgctcag	cgggtgttgg	cgggtgtcgg	ggcgagacca	3960
tgaccagctc	acgtagcgat	agcggagtgt	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcgggtg	tgaataaccg	cacagatgcg	taaggagaaa	4080
ataccgcata	aggcgctctt	ccgcttccct	gctcactgac	tcgctgcgct	cggtcgttcg	4140
gctgcggcga	gcggtatcag	ctcactcaaa	ggcggttaata	cggttatcca	cagaatcagg	4200
ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcggtg	ctggcggttt	tccataggct	ccgcctccct	gacgagcatc	acaaaaatcg	4320
acgtccaagt	cagaggtggc	gaaacccgac	aggactataa	agataaccagg	cgtttcccc	4380
tggaagctcc	ctcgtgcgct	ctcctgttcc	gaccctgccg	cttaccggat	acctgtccgc	4440
ctttctccct	tcgggaagcg	tggcgcttcc	tcatagctca	cgctgtaggt	atctcagttc	4500
gggtgtaggt	gttcgctcca	agctgggctg	tgtgcacgaa	cccccgcttc	agcccgaccg	4560
ctgcgcctta	tccggtaact	atcgtcttga	gtccaaccgc	gtaagacacg	acttatcgcc	4620
actggcagca	gccaactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tggtggccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4740
tctgctgaag	ccagttacct	tcggaaaaag	agttggtagc	tcttgatccg	gcaaacaaac	4800
caccgctggg	agcgggtgggt	tttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gctcgtttga	tctttctgac	ggggtctgac	gctcagtgga	acgaaaactc	4920
acgttaagggt	atgttggtca	tgcattctag	gtactaaaac	aattcatcca	gtaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	5040
ctgttcttcc	ccgatccct	cctgatcga	cgggacgcag	aaggcaatgt	cataccactt	5100
gtccgccttg	ccgcttctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgctg	tctcccaggt	cgcgctggga	aaagacaagt	tcctcttcgg	gcttttccgt	5220
ctttaaaaaa	tcatacagct	cgcgcggatc	tttaaatgga	gtgtcttctt	ccagtttcc	5280
gcaatccaca	tcggccagat	cgttatttcag	taagtaatcc	aattcggcta	agcggctgtc	5340
taagctatcc	gtatagggac	aatccgatat	gtcgatggag	tgaaagagcc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggtc	ttgttcactc	tcatactctt	ccgagcaaac	5460
gacgccatcg	gcctcactca	tgagcagatt	gctccagcca	tcatgccgtt	caaagtgcag	5520
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5580
atcatagggt	gtccctttat	accggctgtc	cgctattttt	aaatatagggt	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgcagcggta	5700
ttttctgato	agttttttca	attccgggtga	tattctcatt	ttagccattt	attatttctc	5760
tctcttttcc	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tccttgcat	ctaaaacctt	aaataaccaga	aaacagcttt	ttcaaaagt	5880
ttttcaaagt	tgccgtataa	catagttatcg	acggagccga	ttttgaaacc	gcggtgatca	5940
caggcagcaa	cgcctgtgca	tcgttacaat	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaaac	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaac	6060
tctgccgcct	tacaacggct	ctcccgtgta	cgcgcgtccc	gactgatggg	ctgcctgtat	6120
cgagtgggtga	ttttgtgccg	agctgcgggt	cggggagctg	ttggctggct	gggtggcagg	6180
tattttgtgg	tgtaaacaaa	ttgacgctta	gacaaactaa	taacacattg	cggagctttt	6240
taattgtactg	aattaacgcc	gaattaatcc	gggggatctg	gatttttagta	ctggattttg	6300
gtttttaggaa	ttagaaattt	tattgataga	agttattttac	aaatacaaat	acatactaag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctatagg	aacctaat	cccttatctg	6420

ggaactactc	acacattatt	atggagaaac	tcgagtcaaa	tctcgggtgac	gggcaggacc	6480
ggacggggcg	gtaccggcag	gctgaagtcc	agctgccaga	aaccacagtc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgccctcgtc	6600
atgcgcacgc	tcgggtcggt	gggcagcccc	atgacagcga	ccacgctctt	gaagccctgt	6660
gcctccaggg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtccegt	ccgctggtgg	6720
cggggggaga	cgtacacggg	cgactcggcc	gtccagtcgt	aggcgttgcg	tgccctccag	6780
gggcccgcgt	aggcgatgcc	ggcgaacctc	ccgtccacct	cggcgacgag	ccagggatag	6840
cgctcccgcg	gacggacgag	gtcgtccgtc	cactcctgcg	gttccctgcg	ctcggtacgg	6900
aagttgaccg	tgcttgcttc	gatgtagtgg	ttgacgatgg	tgacagaccg	cggcattgcc	6960
gcctcgggtg	cacggcggat	gtcggccggg	cgctcgttct	ggctcatggt	agactcgaga	7020
gagatagatt	tgtagagaga	gactgggtgat	ttcagcgtgt	cctctccaaa	tgaatgaac	7080
ttccttatat	agaggaaggt	cttgcggaagg	atagtgggat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaaagt	cttctttttc	7200
cacgatgtcc	ctcgtgggtg	gggggtccatc	tttgggacca	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttcctttat	cgcaatgatg	gcattttgtag	gtgccacctt	ccctttctac	7320
tgtccttttt	atgaagtgc	agatagctgg	gcaatggaat	ccgaggaggt	ttcccgatat	7380
tacccttttg	tgaaaagtct	caatagccct	ttgggtcttct	gagactgtat	ctttgatatt	7440
cttggagtag	acgagagtgt	cggtgtccac	catgtctctc	catcaatcca	cttgctttga	7500
agacgtgggt	ggaaagtctt	ctttttccac	gatgtctctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcttgaac	gatagccctt	cctttatcgc	aatgatggca	7620
tttgtagggt	ccaccttctt	tttctactgt	ccttttctgt	aaagtgcaga	tagctgggca	7680
atggaatccg	aggaggtttc	ccgatattac	cctttgttga	aaagtctcaa	tagccctttg	7740
gtcttctgag	actgtatctt	tgatattctt	ggagtagacg	agagtgtcgt	gctccaccat	7800
gttggcaagc	tgctctagcc	aatacgc aaa	ccgcctctcc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcacgacag	gtttcccgac	tggaaagcgg	gcagtgcgag	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcacc	caggctttac	actttatgct	tccggctcgt	7980
atgttgtgtg	gaatttgtgag	cggataacaa	tttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agccttgact	agagggctga	cggatatacag	acatgataag	atacattgat	8100
gagtttggac	aaaccacaac	tagaatgcag	tgaaaaaaat	gctttatttg	tgaattttgt	8160
gatgtctatt	cttttatttg	aaccattata	agctgcaata	aacaagttgg	ggtgggcgaa	8220
gaactccagc	atgagatccc	cgcgtggag	gatcatccag	ccggcgctcc	ggaaaacgat	8280
tccgaagccc	aacctttcat	agaaggcggc	ggtggaatcg	aaatctcgta	gcacgtgtca	8340
gtcctgctcc	tcggccacga	agtgcacgca	gttgccggcc	gggtcgcgca	gggcgaactc	8400
ccgccccac	cgctcgtcgc	cgatctcggt	catggccggc	ccggaggcgt	cccggaagtt	8460
cgtggacacg	acctccgacc	actcggcgta	cagctcgtcc	aggccgcgca	cccacaccca	8520
ggccagggtg	ttgtccggca	ccacctgggt	ctggaccgcg	ctgatgaaca	gggtcacgtc	8580
gtccccgacc	acaccggcga	agtcgtcttc	acgaaagtcc	cgggagaacc	cgagccggtc	8640
ggtccagaac	tcgaccgctc	cggcgacgtc	gcgcgcgggt	agcaccggaa	cggcactggg	8700
caacttggcc	atggatccag	atttcgctac	agttagtata	aaaaagcagg	cttcaatcct	8760
gcaggaattc	atcgacactg	ctcgtctact	ccaagaatat	caaagataca	gtctcagaag	8820
accaaagggc	tattgagact	tttcaacaaa	gggtaatatc	gggaaacctc	ctcggattcc	8880
attgcccagc	tatctgtcac	ttcatcaaaa	ggacagtaga	aaaggaaggt	ggcacctaca	8940
aatgcccata	ttcgcataaa	ggaaaggcta	tcgttcaaga	tgccctctgc	gacagtggtc	9000
ccaaagatgg	acccccaccc	acgaggagca	tcgtggaaaa	agaagacgtt	ccaaccacgt	9060
cttcaaagca	agtggattga	tgtgataaca	tggtggagca	cgacactctc	gtctactcca	9120
agaatatcaa	agatacagtc	tcagaagacc	aaagggttat	tgagactttt	caacaaaggg	9180
taatatcggt	aaacctcctc	ggattccatt	gccagctat	ctgtcacttc	atcaaaagga	9240
cagtagaaaa	ggaaggtggc	acctacaaat	gccatcattg	cgataaagga	aaggctatcg	9300
ttcaagatgc	ctctgccgac	agtgggtcca	aagatggacc	cccacccacg	aggagcatcg	9360
tggaaaaaga	agacgttcca	accacgtctt	caaagcaagt	ggattgatgt	gatatctcca	9420
ctgacgtaag	ggatgacgca	caatcccact	atccttcgca	agaccttctt	ctatataagg	9480
aagttcattt	catttggaga	ggacacgctg	aaatcaccag	tctctctcta	caaactctatc	9540
tctctcgagc	tttcgcagat	ccgggggggg	aatgagatat	gaaaaagcct	gaactcaccg	9600
cgacgtctgt	cgagaagttt	ctgatcgaaa	agttcgacag	cgtctccgac	ctgatgcagc	9660
tctcggaggg	cgaagaatct	cgtgctttca	gcttcgatgt	aggaggcggt	ggatatgtcc	9720
tgccgggtaaa	tagctgcgcc	gatgggtttct	acaaagatcg	ttatgtttat	cggcactttg	9780
cctcggccgc	gctcccagatt	ccggaaagtgc	ttgacattgg	ggagtttagc	gagagcctga	9840
ccatttgcct	ctcccgctgt	gcacaggggt	tcacgttgca	agacctgcct	gaaaccgaa	9900
tgcccgctgt	tctacaaccg	gtcgcggagg	ctatggatgc	gatcgctgcg	gccgatctta	9960
gccagacgag	cgggttcggc	ccatttcggg	cgcaagggaat	cggtcaatac	actacatggc	10020
gtgatttcat	atgcgcgatt	gctgatcccc	atgtgtatca	ctggcaaaact	gtgatggagc	10080
acaccgtcag	tgcgtccgtc	gcgcaggctc	tcgatgagct	gatgctttgg	gccgaggact	10140
gccccgaagt	ccggcacctc	gtgcacgcgg	atctcggctc	caacaatgtc	ctgacggaca	10200
atggccgcat	aacagcggtc	attgactgga	gcgagggcat	gttcggggat	tcccaatacg	10260
aggtcgcca	catcttcttc	tggaggccgt	ggttggtctg	tatggagcag	cagacgcgct	10320
acttcgagcg	gaggcatccg	gagcttgcat	gatcgccacg	actccggcgc	tatatgctcc	10380
gcattgggtct	tgaccaactc	tatcagagct	tgggtgacgg	caatttcgat	gatgcagctt	10440

-47-

ggggcgagg	tcgatgagac	gcaatcgccc	gatccggagc	cgggactgtc	gggagtagac	10500
aaatcgccc	cagaagcgcg	gccgtctgga	ccgatggctg	tgtagaagta	ctcgccgata	10560
gtggaaaccg	acgccccagc	actcgccgga	gggcaaagaa	atagagtaga	tgccgaccgg	10620
atctgtcgat	cgacaagctc	gagtttctcc	ataataatgt	gtgagtagtt	cccagataag	10680
ggaattaggg	ttcctatagg	gtttcgctca	tgtgttgagc	atataagaaa	cccttagtat	10740
gtatttgtat	ttgtaaaata	cttctatcaa	taaaatttct	aattcctaaa	acaaaaatcc	10800
agtactaaaa	tccagatccc	ccgaattaat	tcggcggttaa	ttcagatcaa	gcttggcact	10860
ggcgctcggt	ttacaacgtc	gtgactggga	aaaccttggc	gttaccacac	ttaatcgctc	10920
tgacgacat	ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	ccgatcgccc	10980
ttcccaacag	ttgcgagcc	tgaatggcga	atgctagagc	agcttgagct	tggtacagat	11040
tgctgtttcc	cgccttcagt	ttaaactatc	agtgtttgac	aggatatatt	ggcggtgtaa	11100
cctaagagaa	aagagcggtt	attagaataa	cggatattta	aaagggcggtg	aaaaggttta	11160
tccgttcgtc	catttgtatg	tg				11182

&lt;210&gt; 90

&lt;211&gt; 8428

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pCambia3300 Plasmid

&lt;400&gt; 90

catgccaaac	acagggttcc	cctcgggagc	aaagtaacttt	gatccaaccc	ctccgctgct	60
atagtgcagt	cggcttctga	cggttcagtgc	agccgtcttc	tgaaaacgac	atgtcgcaca	120
agtcctaagt	tacgcgacag	gctgcccgc	tgcctttttc	ctggcggttt	cttgtcgcgt	180
gttttagtgc	cataaagtag	aatacttgcg	actagaaccg	gagacattac	gccatgaaca	240
agagcgccgc	cgctggcctg	ctgggctatg	cccgcgtcag	caccgacgac	caggacttga	300
ccaaccaacg	ggcggaaactg	cacgcggccg	gctgcaccaa	gctgttttcc	gagaagatca	360
ccggcaccag	gcgcgacccg	ccggagctgg	ccaggatgct	tgaccaccta	cgccctggcg	420
acgttgtgac	agtgaccagg	ctagaccgcc	tgcccgccag	caccgcgcag	ctactggaca	480
ttgcccagcg	catccaggag	gcggcgccgg	gcctgcgtag	cctggcagag	ccgtgggccc	540
acaccaccac	gccggccggc	cgcattggtgt	tgaccgtgtt	cgccggcatt	gcccgttcg	600
agcgttccct	aatcatcgac	cgcaccggga	gcgggcccga	ggccgccaag	gcccagggcg	660
tgaagtttgg	cccccgccct	accctcacc	cggcacagat	cgccgacgcc	cgcgagctga	720
tcgaccagga	aggccgcacc	gtgaaagagg	cggctgcact	gcttggcgtg	catcgctcga	780
ccctgtaccg	cgcacttgag	cgcagcgagg	aagtgcagcc	caccgaggcc	aggcgccgcg	840
gtgccttccg	tgaggacgca	tgaccgagg	ccgacgccct	ggcgccgccc	gagaatgaac	900
gccaaagagga	acaagcatga	aaccgcacca	ggagggccag	gacgaaccgt	ttttcattac	960
cgaagagatc	gaggcgagga	tgatcgccgg	cgggtacgtg	ttcgagccgc	ccgcgcacgt	1020
ctcaaccgtg	cggctgcatg	aaatcctggc	cggtttgtct	gatgccaaag	tgccggccgt	1080
gccggccagc	ttggccgctg	aagaaaccga	gcgcgcgcgt	ctaaaaaggt	gatgtgtatt	1140
tgagtaaaac	agcttgcgtc	atgcggtcgc	tgcttatatg	atgcgatgag	taataaaca	1200
aatacgcaag	gggaacgcat	gaaggttatc	gctgtactta	accagaaagg	cgggtcaggc	1260
aagacgacca	tcgcaaccac	tctagcccgc	gccctgcaac	tcgcccgggg	cgatgttctg	1320
ttagtgcatt	ccgatcccca	gggcagtggc	cgcgattggg	cggccgtgcg	ggaagatcaa	1380
ccgctaaccg	ttgtcggcat	cgaccgccc	acgattgacc	gcgacgtgaa	ggccatcgcc	1440
cggcgcgact	tcgtagtgat	cgacggagcg	cccaggcg	cggacttggc	tgtgtccgcg	1500
atcaaggcag	ccgacttcgt	gctgattccg	gtgcagccaa	gcccttacga	catatgggcc	1560
accgcccagc	tggtggagct	ggttaagcag	ogcattgagg	tcacggatgg	aaggctacaa	1620
gcggcctttg	tcgtgtcgcg	ggcgatcaaa	ggcagcgca	tcggcggtga	ggttgccgag	1680
gcgctggccg	ggtagagct	gcccattctt	gagtcgccga	tcacgcagcg	cgtgagctac	1740
ccaggcactg	ccgcgcgcgg	cacaaccggt	cttgaatcag	aaccgcaggg	cgacgctgcc	1800
cgcgaggtcc	aggcgctggc	cgttgaaatt	aaatcaaaac	tcatttgagt	taatgaggtg	1860
aagagaaaat	gagcaaaagc	acaaacacgc	taagtgccgg	ccgtccgagc	gcacgcagca	1920
gcaaggctgc	aacgttggcc	agcctggcag	acacgccagc	catgaagcgg	gtcaactttc	1980
agttgcccgc	ggaggatcac	accaagctga	agatgtacgc	ggtacgcca	ggcaagacca	2040
ttaccgagct	gctatctgaa	tacatcgccg	agctaccaga	gtaaatgagc	aaatgaataa	2100
atgagtagat	gaatttttagc	ggctaaagga	ggcggtatgg	aaaatcaaga	acaaccaggc	2160
accgacgcgg	tggaaatgcc	catgtgtgga	ggaacgggcg	gttggccagg	cgtaagcgga	2220
tgggttgtct	ccggccctg	caatggcact	ggaaccccca	agcccgagga	atcgccgtga	2280
cggtcgcaaa	ccatccggcc	cggtagcaaat	cggcgccggc	ctgggtgatg	acctggtgga	2340
gaagttgaag	ccgcgcgagg	ccgcccagcg	gcaacgcac	gaggcagaag	cacgcccgcg	2400
tgaatcgtgg	caacggccg	ctgatcgaat	ccgcaagaa	tcccggaac	cgccggcagc	2460
cgggtgcgcg	tcgattagga	agccgcccac	ggcgacgag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tcgcagcatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	ggtgatccgc	tacgagcttc	cagacgggca	2640

cgtagaggtt	tccgcagggc	cggccggcat	ggccagtggt	tgggattacg	acctgggtact	2700
gatggcggtt	tcccatctaa	cogaatccat	gaaccgatac	cgggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaaag	cagaaagacg	acctggtaga	aacctgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggccgcctg	gtgacgggtat	ccgaggggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaacc	gggcggcccg	agtacatcga	3000
gacgagctta	gctgatttga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggacgtgct	3060
gacgggttcac	cccgattact	ttttgatcga	tcccggcatc	ggccggttttc	tctaccgect	3120
ggcacgcgcg	gcccgaggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gccggagagt	tcaagaagtt	ctgtttcacc	gtgcgcaagc	tgatcgggtc	3240
aaatgacctg	ccggagtagc	atttgaagga	ggaggcgggg	caggctggcc	cgatcctagt	3300
catgcgtac	caaattgccc	tcgagggcga	agcatccgcc	ggttcctaata	gtacggagca	3360
gatgctaggg	attgggaacc	tagcagggga	aaaaggctcga	aaaggctctct	ttcctgtgga	3420
tagcacgtac	tacattggga	caaagccgta	cattgggaac	cggaaaccgt	acattgggaa	3480
cccaaagccg	tccgcctaaa	accggtcaca	catgttaagt	actgatataa	aagagaaaaa	3540
aggcgatttt	ctgtctggcc	actctttaaa	acttattaaa	actcttaaaa	cccgcctggc	3600
ctgtgcataa	ccgcctggcc	agcgcacagc	cgaagagctg	caaaaagcgc	ctacccttcg	3660
gtcgctgcgc	ggcctacggc	ccgcctgttc	gcgtcggcct	atcgccggccg	ctggccgctc	3720
aaaaatggct	cgccgcccac	caggcaatct	accagggcgc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	aaaacctctg	atcaaggcac	cctgcctcgc	gcgtttcggg	gatgacgggt	3840
aaaacctctg	agcccgtag	ctcccggaga	cgggtgttgg	ttgtctgtaa	gcggatcccg	3900
ggagcagaca	acgtagcgat	ggcgcgtag	atactggctt	cgggtgtcgg	ggcgagcca	3960
tgaccagtc	agagtgacac	agcggaggtg	tgaaataccg	aactatgcgg	catcagagca	4020
gattgtactg	aggcgctctt	atattgcggtg	gctcactgac	cacagatgcg	taaggagaaa	4080
ataccgcac	gcggtatcag	ccgcttcttc	ggcggtataa	tcgctgcgct	cggtcgttcg	4140
gctgcggcga	ggaaagaaca	ctcactcaaa	agggccagca	cggttatcca	cagaatcagg	4200
ggataacgcg	ctggcggttt	tgtgagcaaa	ccgccccctt	aaggccagga	accgtaaaaa	4260
ggccgcggtg	cagaggtggc	tccataggct	aggactataa	gacgagcatc	acaaaaatcg	4320
acgctcaagt	ctcgtgcgct	gaaacccgac	gacctgcgcg	agataccagg	cgtttccccc	4380
tggaagctcc	tcgggaagcg	ctcctgttcc	tcatagctca	cttaccggat	acctgtccgc	4440
ctttctccct	gttcgctcca	tggtgcggtg	tgtgcacgaa	cgctgtaggt	atctcagttc	4500
gggtgtaggt	tcgggtaact	atcgtcttga	gtccaacccg	ccccccgttc	agcccgacgc	4560
ctgcgcctta	gccactggta	acaggattag	cagagcgagg	gtaagacacg	acttatcgcc	4620
actggcagca	tgttcttgaag	actacggcta	cactagaagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tcggggcccta	tcggaaaaag	agttgtagc	acagtatttg	gtatctgcgc	4740
ctctgtgaag	ccagttacct	tttttgtttg	caagcagcag	tcttgatccg	gcaaacaaa	4800
caccgctggg	agcgggtgggt	tcttttctac	ggggtctgac	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gatcccttga	tgcatcttag	ggggtctgac	gctcagtgga	acgaaaaact	4920
acgttaaggg	atcttgggtca	caggcttgat	aaacataaac	aattcatcca	gtaaaaatata	4980
atattttatt	ttctcccaat	ccctgatcga	cccagtaag	tcaaaaaata	gctcgacata	5040
ctgttcttcc	ccgatattct	caagatcaat	ccggagcgag	aaggcaatgt	cataccactt	5100
gtccgcctcg	ccgcttctcc	caagcactt	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgcgt	tctcccaggt	aaagacaagt	aaagacaagt	tcctcttcgg	gcttttccgt	5220
ctttaaaaaa	tcatacagct	tttaaatgga	tttaaatgga	gtgtcttctt	cccagttttc	5280
gcaatccaca	tcggccagat	taagtaattc	taagtaattc	aattcggcta	agcggctgtc	5340
taagctatct	gtatagggac	aatccgatat	gtcgatggag	tgaaagagcc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggct	ttgttcatct	tcatactctt	ccgagcaaac	5460
gacgccatcg	gcctcactca	tgagcagatt	gctccagcca	tcagtcggtt	caaagtgcag	5520
gaccttttga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5580
atcataggtg	gtccctttat	accggtgtgc	cgtcattttt	aaatataggt	tttcattttc	5640
ttccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgcagcggta	5700
tttttcgact	agttttttca	attccggtga	tattctcatt	ttagccattt	attatttctt	5760
tcctcttttc	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tccttgcat	ctaaaacctt	aaataccaga	aaacagcttt	ttcaaagttg	5880
ttttcaaagt	tggcgataaa	catagtatcg	acggagccga	ttttgaaacc	gcgggtgatca	5940
caggcagcaa	cgctctgtca	tcgttacaa	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaacc	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaac	6060
ctctccgctc	tacaacggct	ctcccgtgta	cgccgtcccg	gactgatggg	ctgcctgtat	6120
cgagtgggtga	ttttgtgccc	agctgccggg	cggggagctg	ttggctggct	gggtggcagga	6180
tatatgttgg	tgtaaacaaa	ttgacgctta	gacaacttaa	taacacattg	cggacgtttt	6240
taattactcg	gaattaatcc	gaattaatcc	gggggatctg	gatttttagta	ctggattttg	6300
gttttaggaa	ttagaaattt	tattgataga	agttattttac	aaatacaaat	acataactaag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctatagg	aaccttaatt	cccttatctg	6420
ggaactactc	acacattatt	atggagaaac	tcctggtgac	tctcggtgac	gggcaggacc	6480
ggacggggcg	gtaccggcag	gctgaagtcc	agctgccaga	aaccacacgc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgccctcgtc	6600
atgcgcacgc	tcgggtcggt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660



-49-

```

gcctccaggg acttcagcag gtgggtgtag agcgtggagc ccagtcctcgt ccgctgggtgg 6720
cggggggaga cgtacacggg cgactcggcc gtccagtcgt aggcgttgcg tgccttccag 6780
ggggcccgct aggcgatgcc ggccaccccg cgtccacct cggcgacgag ccagggatag 6840
cgctcccgca gacggacgag gtcgtccgtc cactccctcg gttccctgcgg ctccggtacgg 6900
aagttgaccg tgcttgcttc ttgacgatgg ttgacgatgg tgcagaccgc cggcatgtcc 6960
gcctcgggtg cacggcggat gtccggcggg cgtcgttctg ggctcatggg agactcgaga 7020
gagatagatt tgtagagaga gactgggtgat ttcagcgtgt cctctccaaa tgaaatgaac 7080
ttccttatat tcacatcaat cctgcgaagg ccacttgctt atagtgggat cccttacgtc 7140
agtggagata ctcgtgggtg ggggtccatc tgaagacgtg gttggaacgt cttctttttc 7200
cacgatgctc tttcctttat tttgggacca ctgtcggcag aggcattctt 7260
aacgatagcc tttcctttat gcatctgtag gcaatggaaat gtgccacctt ccttttctac 7320
tgctcctttg atgaagtac agatagctgg gcaatggaaat cggaggaggt tccccgatat 7380
taccctttgt tgaaaagtct caatagccct ttggtcttct gagactgtat ctttgatatt 7440
cttggagtag acgagagtgt cgtgctccac catgttatca catcaatcca cttgctttga 7500
agacgtgggt ggaacgtctt gatgctcctc gtaggccttt gtgggtgggg gtcctatctt 7560
gggaccactg tcggcagagg catcttgaac gatagccctt cctttatcgc aatgatggca 7620
ttttagtggt ccaccttctt ccttttgatg aagtgcacaga tagctgggca 7680
atggaatccg aggaggttct ccttttgatg aaagtctcaa tagccctttg 7740
gtcttctgag actgtatctt tgatattctt ggagtagacg agagtgctgt gctccaccat 7800
gttggcaagc tgctctagcc aatacgcaaa ccgcctctcc ccgcgcgttg gccgattcat 7860
taatgcagct ggcacgacag gtttcccgac gtagagctcg acctgcaggc caacgcaatt 7920
aatgtgagtt agctcactca ttaggcacc caggctttac actttatgct tccggctcgt 7980
atgttgtgtg gaattgtgag cggggaatcc tttcacacag gaaacagcta tgaccatgat 8040
tacgaattcg agctcggtag cgggggatcc tctagagtcg acctgcaggc atgcaagctt 8100
ggcactggcc gtcgttttac aacgtcgtga ctgggaaaac cctggcggtta cccaacttaa 8160
tcgccttgca gcacatcccc ctttcgccag ctggcgtaat agcgaagagg cccgcaccga 8220
tcgcccctcc caacagttgc gcagcctgaa tggcgaatgc tagagcagct tgagcttggg 8280
tcagattgtc gtttcccgcc ttccagttta actatcagtg tttgacagga tatattggcg 8340
ggtaaaccta agagaaaaga gcgtttatta gaataacgga tatttaaaag ggcgtgaaaa 8400
ggtttatccg ttcgtccatt tgtatgtg 8428

```

&lt;210&gt; 91

&lt;211&gt; 3438

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pLIT38attBZeo Plasmid

&lt;400&gt; 91

```

tcgaccctct agtcaaggcc ttaagttagt cgtattacgg actggcgcgtc gttttacaac 60
gtcgtgactg ggaataaccct ggcggttacc aacttaatcg ccttgacgca catccccctt 120
tcgccagctg gcgtaaatagc gaagaggccc gcaccgatcg ccttcccaa cagttgcgca 180
gcctgaatgg cgaatggcgc ttccgcttgg aataaagccc gcttcggcgg gctttttttt 240
gttaactacg tcaggtggca cttttcgggg aaatgtgcgc tttgtttatt 300
tttctaaata cattcaaata tgtatccgct catgagacaa taaccctgat aaatgcttca 360
ataatattga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgcgc ttattccctt 420
ttttgcggca ttttgccctc ctgtttttgc tcaccagaa acgctggtga aagtaaaaga 480
tgctgaagat cagttgggtg cacgagtggg ttacatcgaa ctggatctca acagcggtaa 540
gatccttgag agttttcgcc ccgaagaacg ttctccaatg atgagcactt ttaaagtctt 600
gctatgtggc gcggtattat cccgtgttga cgccgggcaa gagcaactcg gtcgccgcat 660
acactattct cagaatgact tggttgagta ctcaccagtc acagaaaagc atcttacgga 720
tgccatgaca gtaagagaat tatgcagtgc tgccataacc atgagtata acactgcggc 780
caacttactt ctgacaacga tcggaggacc gaaggagcta accgcttttt tgcacaacat 840
gggggatcat gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag ccataccaaa 900
cgacgagcgt gacaccacga tgctgtagc aatggcaaca acgttgcgca aactattaac 960
tgccgaacta cttactctag cttcccgcca acaattaata gactggatgg aggcggataa 1020
agttgcagga ccaactctgc gctcggctggc tccggctggc tggtttattg ctgataaact 1080
tgagcgggt gagcgtgggt ctccggttat cattgcagca ctggggccag atggtaagcc 1140
ctcccgatc gatagttatc acacgacggg gactcaggca actatggatg aacgaaatag 1200
acagatcgct cctcactagg cctcactagg taagcattgg taactgtcag accaagttta 1260
ctcatatata ctttagattg atttaccctg gttgataatc agaaaagccc caaaaacagg 1320
aagattgtat aagcaaatat ttaaattgta aacgttaata ttttgtaaaa attcgcgtta 1380
aatttttgtt aattcagctc caataggccc aaatcggcaa aatcccttat 1440
aatcaaaaag aatagcccgga agtgttgttc cagtttggaa caagagtcca 1500
ctattaaaga acgtggactc caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc 1560
ccactacgtg aaccatcacc caaatcaagt tttttggggg cgaggtgcgc taaagcacta 1620

```

-50-

```
aatcggaaacc ctaaagggag cccccgattt agagcttgac ggggaaagcg aacgtggcga 1680
gaaagggaagg gaagaaagcg aaaggagcgg gcgctagggc gctggcaagt gtagcggtca 1740
cgctgcgcgt aaccaccaca cccgcgcgcg ttaatgcgcg gctacagggc gcgtaaaagg 1800
atctaggtga agatcctttt tgataatctc atgacaaaaa tcccttaacg tgagttttcg 1860
ttccactgag cgtcagaccc cgtagaaaag atcaaaggat cttcttgaga tccttttttt 1920
ctgcgcgtaa tctgctgctt gcaaacaaaa aaaccaccgc taccagcggg ggtttggttg 1980
ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata 2040
ccaaatactg ttcttctagt gtagccgtag ttaggccaac acttcaagaa ctctgtagca 2100
ccgcctacat acctcgctct gctaactctg ttaccagtgg ctgctgccag tggcgataag 2160
tcgtgtctta ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc 2220
tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 2280
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aaggagagaa ggccgacagg 2340
tatccggtaa gcggcagggg cggaaacagga gagcgcacga gggagcttcc aggggggaaac 2400
gcctgggtac ttctatagtc cgtcgggttt cgccacctct gacttgagcg tcgatttttg 2460
tgatgctcgt cagggggggc gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg 2520
ttcctggcct tttgctggcc ttttgctcac atgtaatgtg agttagctca ctcattaggc 2580
acccaggtt ttacaactta tgcttccggc tcgtatgttg tgtggaattg tgagcggata 2640
acaatttcac acaggaacaa gctatgacca tgattacgcc aagctacgta atacgactca 2700
ctagtggggc ccgtgcaatt gaagccggct ggccccaagc ttctctgcag gattgaagcc 2760
tgctttttta tactaacttg agcgaatct ggatccatgg ccaagttagc cagtgcctgt 2820
ccgggtgctc ccgcgcgga cgtcgccgga gcggtcgagt tctggaccga ccggctcggg 2880
ttctcccggg acttcgtgga ggacgacttc gccggtgtgg tccgggacga cgtgaccctg 2940
ttcatcagcg cgggtccagga ccaggtggtg ccggacaaca ccctggcctg ggtgtgggtg 3000
cgcgccctgg acgagctgta ccgcgagtg cgcgaggtcg tgtccacgaa cttccgggac 3060
gcctccgggc cggccatgac cgagatcggc gagcagccgt gggggcgggg gttcgccctg 3120
cgcgaccggc ccggcaactg cgtgcacttc gtggccgagg agcaggactg acacgtgcta 3180
cgagatttcg attccacgc cgcttctat gaaaggttgg gcttcggaat cgtttccgg 3240
gacgcgggct ggatgatcct ccagcgcggg gatctcatgc tggagttctt cgcccccccc 3300
aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 3360
aataaagcat tttttcact gcattctagt tgtggtttgt ccaaactcat caatgtatct 3420
tatcatgtct gtataccg 3438
```

<210> 92  
<211> 10549  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> pCambia1302 Plasmid

<300>  
<308> Genbank #AF234398  
<309> 2000-04-24

```
<400> 92
catggtagat ctgactagta aaggagaaga acttttctact ggagttgtcc caattcttgt 60
tgaattagat ggtgatgtta atgggcacaa attttctgtc agtggagagg gtgaaggtga 120
tgcaacatac ggaaaactta cccttaaatt tatttgcact actggaaaac tacctgttcc 180
gtggccaaca cttgtcacta ctttctctta tgggtgttcaa tgcttttcaa gatacccaga 240
tcatatgaag cggcacgact tcttcaagag cgccatgcct gagggatagc tgcaggagag 300
gaccatcttc ttcaaggacg acgggaacta caagacaggt gctgaagtca agtttgaggg 360
agacaccctc gtcaacagga tcgagcttaa gggaatcgat ttcaaggagg acggaaacat 420
cctcggccac aagttggaat acaactacaa ctcccacaa gtatacatca tggccgacaa 480
gcaaaagaac ggcatacaa gccaactcaa gacccggcac aacatcgaag acggcgggcg 540
gcaactcgct gatcattatc aacaaaatac tccaattggc gatggccctg tccttttacc 600
agacaacctat tacctgtcca cacaatctgc cctttcgaaa gatcccaacg aaaagagaga 660
ccacatggtc cttcttgagt ttgtaacagc tgctgggatt acacatggca tggatgaact 720
atacaaagct agccaccacc accaccacca cgtgtgaatt ggtgaccagc tcgaatttcc 780
ccgatcggtt aaacatttgg caataaagtt tcttaagatt gaatcctgtt gccgggtctg 840
cgatgatttc catataattt ctgttgaatt agttaagca tgtaataatt aacatgtaact 900
gcatgacgtt atttatgaga tgggttttta tgattagagt cccgcaatta tacatttaat 960
acgcgataga aaacaaaata tagcgcgcaa actaggataa attatcgcg gcgggtgtcat 1020
ctatgtttat agatcgggaa ttaactatc agtgtttgac aggatataat ggcgggtaaa 1080
cctaagagaa aagagcggtt attagaataa cggatattta aaagggcggtg aaaaggttta 1140
tccgttcgtc catttgatat tgcattgcaa ccacaggggt cccctcggga tcaaagtaact 1200
ttgatccaac ccctccgctg ctatagtgca gtgcggtctt gacgttcagt gcagcgtct 1260
tctgaaaacg acatgtcgca caagtccata gttacgcgac aggtgcgcgc cctgcccttt 1320
```



tccctggcggtt	ttcttgtgcgc	gtgttttagt	cgcataaagt	agaataacttg	cgactagaac	1380
cggagacatt	acgccatgaa	caagagcgcc	gccgctggcc	tgctgggcta	tgcccgcgtc	1440
agcaccgacg	accaggactt	gaccaacca	cgggccgaac	tgacgcggc	cggtgcacc	1500
aagctgtttt	ccgagaagat	caccggcacc	aggcgcgacc	gcccggagct	ggccaggatg	1560
cttgaccacc	tacgccctgg	cgacgttggt	acagtgacca	ggctagaccg	cctggccccg	1620
agcaccgcg	acctactgga	cattgccgag	cgcattccagg	aggccggcgc	gggcctgcgt	1680
agcctggcag	agccgtgggc	cgacaccacc	acgcccggcg	gccgcatggt	ggtgaccgtg	1740
ttcgccggca	ttgccgagtt	cgagcggttc	ctaatacatcg	accgcaccgc	gagcggggcg	1800
gaggccgcca	aggcccagg	cgtgaagttt	ggcccccgcc	ctaccctcac	cccggcacag	1860
atcgcgacg	cccgcgagct	gatcgaccag	gaaggccgca	ccgtgaaaga	ggcggctgca	1920
ctgcttggcg	tgcatcgctc	gaccctgtac	cgcgcacttg	agcgcagcga	ggaagtgcag	1980
cccaccgagg	ccaggcgggc	cggtgccttc	cgtgaggacg	cattgaccga	ggccgacgcc	2040
ctggcgggcg	ccgagaatga	acgccaagag	gaacaagcat	gaaaccgcac	caggacggcc	2100
aggacgaacc	gtttttcatt	accgaagaga	tcgaggcgga	gatgatcgcg	gcccgggtacg	2160
tggtcgagcc	gcccgcgcac	gtctcaaccg	gtcggtgca	tgaaatcctg	gcccgttctg	2220
ctgatgccaa	gctggcgggc	tgccgggcca	gcttggcgcg	tgaagaaacc	gagcgccgccc	2280
gtctaaaaag	gtgatgtgta	tttgagtaaa	acagcttgcg	tcatgcgggtc	gctgcgtata	2340
tgatgcatg	agtaaaataaa	caaatacgca	agggggaacg	atgaagggtta	tcgctgtact	2400
taaccagaaa	ggcgggtcag	gcaagagcac	cattgcacac	catctagccc	gccccttgca	2460
actcgccggg	gccgatgttc	tgttagtcga	ttccgatccc	cagggcagtg	cccgcgattg	2520
ggcgcccggt	cggggaagatc	aaccgctaac	cgttgtcgcc	atcgaccgcc	cgacgattga	2580
cccgacgctg	aaggccatcg	gccggcgcgga	cttcgtagtg	atcgacggag	cgccccaggc	2640
ggcggaactg	gctgtgtccg	cgatcaaggc	agccgacttc	gtgctgattc	cggtgcagcc	2700
aagcccttac	gacatatggg	ccaccgcgga	cctggtggag	ctggttaagc	agcgattgca	2760
ggtcacggtt	ggaaggctac	aagcggcctt	tgctggtcg	cgggcgatca	aaggcacgcg	2820
catcgccggt	gaggttgccg	aggcgctggc	cggttacgag	ctgcccattc	ttgagtcgcc	2880
tatcacgctg	cgcgtagact	acccaggcac	tgccgcgcgc	ggcacaaccg	ttcttgaatc	2940
agaaccgag	ggcgacgctg	cccgcgaggt	ccaggcgctg	gccgctgaaa	ttaaatcaaa	3000
actcatttga	gttaatgagg	taaagagaaa	atgagcaaaa	gcacaaacac	gctaagtggc	3060
ggcgctccga	ggcgacgcag	cagcaaggct	gcaacgttgg	ccagcctggc	agacacgcca	3120
gccatgaagc	gggtcaactt	tcagttgcog	gcaggaggtc	acaccaagct	gaagatgtac	3180
gcggtacgcc	aaggcaagac	cattaccgag	ctgctatctg	aatacatcgc	gcagctacca	3240
gagtaaatga	gcaaatgaat	aaatgagtag	atgaatttta	gcggttaaag	gaggcggtat	3300
ggaaaatcaa	gaacaaccag	gcaccgacgc	cgtggaatgc	cccatgtgtg	gaggaaacggg	3360
cggttggcca	ggcgtaagcg	gctgggttgt	ctgccggccc	tgcaatggca	ctggaacccc	3420
caagcccgag	gaatcggcgt	gacggtcgca	aaccatccgg	cccggtacaa	atcggcgcgg	3480
cgctgggtga	tgacctgggtg	gagaagttga	aggccgcgca	ggccgcccag	cgccgacgca	3540
tcgaggcaga	agcacgcccc	ggtgaatcgt	ggcaagcggc	cgctgatcga	atccgcaaa	3600
aataccggga	accccgggca	gccggtgcgc	cgtcgattag	gaagccgccc	aagggcgacg	3660
agcaaccaga	tttttctggt	ccgatgctct	atgacgtggg	caccgcgat	agtcgcagca	3720
tcattggacgt	ggccggttttc	cgtctgtcga	agcgtgaccg	acgagctggc	gaggtgatcc	3780
gctacgagct	tcacgacggg	cacgtagagg	tttccgcagg	gcccggccggc	atggccagtg	3840
tgtaggatta	cgacctggta	ctgatggcgg	tttccatct	aaccgaatcc	atgaaacgat	3900
accgggaagg	gaagggagac	aagcccggcc	gcgtgttccg	tcacacagtt	gcggaacgtac	3960
tcaagttctg	ccggcgagcc	gatggcgga	agcagaaaga	cgacctggta	gaaacctgca	4020
ttcgggtaaa	caccacgcac	gttgccatgc	agcgtacgaa	gaaggccaag	aacggccgcc	4080
tggtgacggt	atccgagggg	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	4140
ccgggcggcc	ggagtagatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	4200
aaggcaagaa	cccggacgtg	ctgacggttc	accccgatta	ctttttgata	gatcccggca	4260
tcggccgttt	tctctaccgc	ctggcacgcc	gcgcgcgagg	caaggcagaa	gccagatggg	4320
tggtcaagac	gatctacgaa	cgcatgggca	gcccgggaga	gttcaagaag	ttctgtttca	4380
ccgtgcgcaa	gctgatcggg	tcaaatgacc	tgccggagta	cgatttgaag	gaggaggcgg	4440
ggcaggctgg	ccgatcccta	gtcatgcgct	accgcaacct	gatcgagggc	gaagcatccg	4500
ccggttccta	atgtacggag	cagatgctag	ggcaaatatgc	cctagcaggg	gaaaaagggtc	4560
gaaaagggtc	ctttcctgtg	gatagcgctg	acattgggaa	cccaaagccg	tacatgggga	4620
accggaaccc	gtacattggg	aacccaaagc	cgtacattgg	gaaccgggtca	cacatgtaag	4680
tgactgatat	aaaagagaaa	aaaggcgatt	tttccgccta	aaactcttta	aaacttatta	4740
aaactcttaa	aaccgccttg	gcctgtgcac	aactgtctgg	ccagegcaca	gcccgaagagc	4800
tgcaaaaagc	gcctaccctt	cggtgcgtgc	gctccctacg	ccccgcgcgt	tcgcgtcgcc	4860
ctatcgccgc	cgctggccgc	tcaaaaatgg	ctggcctacg	gccaggcaat	ctaccagggc	4920
gcggacaagc	cgcgccgtcg	ccactcgacc	cgccggcccc	acatcaaggc	acctgcctc	4980
gcgcgtttcg	gtgatgacgg	tgaaaaacctc	tgacacatgc	agctcccggg	gacgggtcaca	5040
gcttgtctgt	aagcggatgc	cgggagcaga	caagcccgtc	agggcgcgctc	agcgggtgtt	5100
ggcgggtgtc	ggggcgccag	catgacccag	tcacgtagcg	atagcggagt	gtatactggc	5160
tttaactatgc	ggcatcagag	cagatgttac	tgagagtgc	ccatatgcgg	tgtgaaatac	5220
cgcacagatg	cgtaaggaga	aaataccgca	tcaggcgctc	ttccgcttcc	tcgctcactg	5280
actcgctgog	ctcgggtcgtt	cggctgcggc	gagcgggtatc	agctcactca	aaggcgggtaa	5340

tacggttatc	cacagaatca	ggggataaac	caggaaagaa	catgtgagca	aaaggccagc	5400
aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggt	tttccatagg	ctccgcccc	5460
ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagagggt	gcgaaacccg	acaggactat	5520
aaagatacca	ggcggtttccc	cctggaagct	ccctcgtgcg	ctctcctggt	ccgaccctgc	5580
cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	cgtggcgctt	tctcatagct	5640
cacgctgtag	gtatctcagt	tcgggtgtagg	tcgttcgctc	caagctgggc	tgtgtgcacg	5700
aaccccccg	tcagcccgac	cgctgcgcct	tatccggtaa	ctatcgtctt	gagtccaacc	5760
cggttaagaca	cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	5820
ggatatgtagg	cggtgctaca	gagttcttga	agtgggtggc	taactacggc	tacactagaa	5880
ggacagtatt	tggtatctgc	gctctgctga	agccagttac	cttcggaaaa	agagtgggta	5940
gctcttgatc	cggcaaacaa	accaccgctg	gtagcgggtg	tttttttgg	tgcaagcagc	6000
agattacg	cagaaaaaaa	ggatctcaag	aagatccttt	gatcttttct	acggggtctg	6060
acgctcagtg	gaacgaaaa	tcacgttaag	ggattttgg	catgcattct	aggtactaaa	6120
acaattcatc	cagtaaaata	taatatctta	ttttctccca	atcaggcttg	atccccagta	6180
agtcaaaaa	tactgttctt	tactgttctt	ccccgatata	ctccctgatc	gaccggacgc	6240
agaaggcaat	gtcataccac	ttgtccgccc	tgccgcttct	cccaagatca	ataaagccac	6300
ttactttg	atctttcaca	aagatgttgc	tgctctcccg	gtcgcgcgtg	gaaaaagacaa	6360
ggtcctcttc	gggcttttcc	gtcttttaaaa	aatcatacag	ctcgcgcgga	cttttaaatg	6420
gagtgtcttc	ttcccagttt	tcgcaatcca	catcggccag	atcgttattc	agtaagtaat	6480
ccaattcg	taagcggctg	tctaaagctat	tcgtataggg	acaatccgat	atgtcgtatg	6540
agtgaagag	ctgatgcac	tcgcataaca	gctcgataat	cttttcaggg	ctttgttctt	6600
cttcatactc	ttccgagcaa	aggacgccat	ggcctcact	catgagcaga	ttgtctccagc	6660
catcatgccc	ttcaaaagtgc	aggacctttg	gaacaggcag	ctttccttcc	agccatagca	6720
tcattgtctt	ttcccgttcc	acatcatagg	tggtcccttt	ataccggctg	tccgtcattt	6780
ttaaatatag	gttttctatt	tctcccacca	ggttatatac	cttagcagga	gacattcctt	6840
ccgtatcttt	tacgcagcgg	tatttttcga	tcagtttttt	caattccggg	gatatttctc	6900
tttttagccat	ttattatttc	cttctctctt	tctacagtat	ttaaagatac	cccaagaagc	6960
taattataac	aagacgaact	ccaattcact	gttctctgca	ttctaaaacc	ttaaatacca	7020
gaaaacagct	ttttcaaagt	tgttttcaaa	gttggcgtat	aacatagtat	cgacggagcc	7080
gattttgaaa	ccgcgggtgat	cacaggcagc	aacgctctgt	catcggtaca	atcaacatga	7140
tacctccgcg	gagatcatcc	gtgtttcaaa	cccgccagct	tagttgccc	tcttccgaat	7200
agcatcggt	acatgagcaa	agtctgccgc	cttacaacgg	ctctcccgc	gacgcgcgtc	7260
cggaatgatg	ggctgcctgt	atcgagtggt	gattttgtgc	cgagctgccg	gtcggggagc	7320
tggtggctgg	ctgggtggcag	gatatattgt	ggtgtaaaaca	aattgacgct	tagacaactt	7380
aataacacat	tgccggacgtt	tttaatgtac	tgaattaaacg	ccgaattaat	tcgggggatc	7440
tggtatttat	tactggattt	tggttttagg	aattagaaat	tttattgata	gaagtatttt	7500
acaaatacaa	atacatacta	agggtttctt	atatgtctcaa	cacatgagcg	aaacctata	7560
ggaaccctaa	ttcccttatc	tggaacttac	tcacacatta	ttatggagaa	actcgagctt	7620
gctgatcgac	agatccggtc	ggcatctact	ctatttcttt	gccctcggac	gagtgtctgg	7680
gcgtcggttt	ccactatcgg	cgagtacttc	tcacagacca	tcgggtccaga	cggtcgctgt	7740
tctgcccggc	atgtgtgtac	gcccagacgt	cccggctccg	gatcggacga	ttgcgtcgca	7800
tcgaccctgc	gcccagctgc	catcatcgaa	attgccgtca	accaagctct	gatagagtgt	7860
gtcaagacca	atcgggagca	tatacgcccg	gagtcgtggc	gatcctgcaa	gctccggatg	7920
cctccgctcg	aagtagcgcg	tctgctgctc	catacaagcc	aaccacggcc	tccagaagaa	7980
gatgttggcg	acctcgat	gggaatcccc	gaacatcgcc	tcgctccagt	caatgaccgc	8040
tggtatgcgg	ccattgtccg	tcaggacatt	gttggagccg	aaatccgcgt	gcacgaggtg	8100
ccggacttcg	gggcagtcct	cggcccaaag	catcagctca	tcgagagcct	gcgcgacgga	8160
cgactgacg	gtgtcgtcca	tcacagtttg	ccagtatac	acatggggat	cagcaatcgc	8220
gcatatgaaa	tcacgccatg	tagtgtattg	accgattcct	tgcggtccga	atgggcccga	8280
cccgcctcgt	tggetaagat	cggccgcagc	gatcgcatcc	atagcctccg	cgaccggttg	8340
tagaacagcg	ggcagttcgg	tttcaggcag	gtcttgcaac	gtgacaccct	gtgcacggcg	8400
ggagatgcaa	taggtcaggg	tctcgctaaa	ctcccgaatg	tcaagcactt	ccggaatcgg	8460
gagcgcggcc	gatgcaaagt	gccgataaac	ataacgatct	ttgtagaaac	catcggcgca	8520
gctatttacc	cgcaggacat	atccacgccc	tcctacatcg	aagctgaaag	cacgagattc	8580
ttcgccctcc	gagagctgca	tcaggtcgga	gacgctgtcg	aacttttctga	tcagaaactt	8640
ctcgacagac	gtcgcggtga	gttcaggctt	tttcatatct	cattgcccc	cgggatctgc	8700
gaaagctcga	gagagataga	tttgtagaga	gagactgggt	atttcagcgt	gtcctctcca	8760
aatgaaatga	acttctctat	atagaggaag	gtcttgcgaa	ggatagtggg	attgtgcgtc	8820
atcccttacg	tcagtggaga	tatcacatca	atccacttgc	tttgaagacg	tggttggaac	8880
gtcttctttt	tccacgatgc	tctctcgtgg	tgggggtcca	tctttgggac	cactgtcggc	8940
agaggtcatc	tgaacgatag	cctttccttt	atcgcaatga	tggcatttgt	aggtgccacc	9000
ttccttttct	actgtccttt	tgatgaagtg	acagatagct	gggcaatgga	atccgaggag	9060
gtttcccgat	attacccttt	gttgaaaagt	ctcaatagcc	ctttgggtctt	ctgagactgt	9120
atctttgata	ttcttggagt	agacgagagt	gtcgtgtctc	accatgttat	cacatcaact	9180
cacttgcttt	gaagacgtgg	ttggaacgtc	ttcttttctc	acgatgctcc	tcgtgggtgg	9240
gggtccatct	ttgggaccac	tgctcggcaga	ggcatcttga	acgatagcct	ttcctttatc	9300
gcaatgatgg	catttgtagg	tgccaccttc	cttttctact	gtccttttga	tgaagtgaca	9360

-53-

```

gatagctggg caatggaatc cgaggagggtt tcccgatatt accctttggtt gaaaagtctc 9420
aatagccctt tgggtcttctg agactgtatc tttgatattc ttggagtaga cgagagtgtc 9480
gtgctccacc atgttggcaa gctgctctag ccaatacgca aaccgcctct ccccgcgct 9540
tggccgattc attaatgcag ctggcacgac aggtttcccg actggaaagc gggcagttag 9600
cgcaacgcaa ttaatgtgag ttagctcact cattaggcac ccagggttt acactttatg 9660
cttcagggtc gtatgttggt tgggaattgtg agcggataac aatttcacac aggaaacagc 9720
tatgaccatg attacgaatt cgagctcggg acccggggat cctctagagt cgacctgcag 9780
gcatgcaagc ttggcactgg ccgtcggttt acaacgtcgt gactgggaaa accctggcgt 9840
taccacaactt aatcgcttg cagcacatcc ccttttcgcc agctggcgta atagcgaaga 9900
ggcccgacc gatcgccctt cccaacagtt gcgcagcctg aatggcgaat gctagagcag 9960
cttgagcttg gatcagattg tcgtttcccg ccttcagttt agcttcatgg agtcaaagat 10020
tcaaatagag gacctaacag aactcgccgt aaagactggc gaacagttca tacagagtct 10080
cttacgactc aatgacaaga agaaaatctt cgtcaacatg gtggagcacg acacacttgt 10140
ctactccaaa aatatcaaa atacagtctc agaagaccaa agggcaattg agacttttca 10200
acaaagggtc atatccggaa acctcctcgg attccattgc ccagctatct gtcactttat 10260
tgtgaagata gtggaaaagg aagggtggctc ctacaaatgc catcattgcg ataaaggaaa 10320
ggccatcggt gaagatgcct ctgccgacag tggccccaaa gatggacccc caccacgag 10380
gagcatcggt gaaaagaag acgttccaac cactgttca aagcaagtgg attgatgtga 10440
tatctccact gacgtaagg atgacgcaca atcccactat ccttcgcaag acccttcctc 10500
tatataagga agttcatttc atttggagag aacacggggg actcttgac 10549

```

<210> 93  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> CaMV35SpolyA Primer

<400> 93  
 ctgaattaac gccgaattaa ttcgggggat ctg 33

<210> 94  
 <211> 29  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> CaMV35Spr Primer

<400> 94  
 ctagagcagc ttgccaacat ggtggagca 29

<210> 95  
 <211> 12592  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pAg2 Plasmid

```

<400> 95
gtacgaagaa ggccaagaac ggccgcctgg tgacgggtatc cgagggtgaa gccttgatta 60
gccgctacaa gatcgtaaa agcgaaccgg ggccggccgga gtacatcgag atcgagctag 120
ctgattggat gtaccgag atcacagaag gcaagaaccc ggacgtgctg acggttcacc 180
ccgattactt tttgatcgat ccgggcacgc gccgttttct ctaccgcctg gcacgcccgc 240
ccgcaggcaa ggcagaagcc agatggttgt tcaagacgat ctacgaacgc agtggcagcg 300
ccggagagtt caagaagttc tgtttcaccg tgccgaagct gatcgggtca aatgacctgc 360
cggagtacga tttgaaggag gaggcggggc aggtcggccc gatcctagtc atgcgctacc 420
gcaacctgat cgagggcgaa gcatccgccc gttcctaatt tacggagcag atgctagggc 480
aaattgccct agcaggggaa aaaggtcgaa aaggctctct tctgtggat agcacgtaca 540
ttgggaaccc aaagccgtac attgggaacc ggaacccgta cattgggaac ccaaagccgt 600
acattgggaa ccggtcacac atgtaagtga ctgatataaa agagaaaaaa ggcgattttt 660
ccgcctaaaa ctttttaaaa ctcttaaaac ccgcctggcc tgtgcataac 720
tgtctggcca gcgcacagcc gaagagctgc aaaaagcgcc tacccttcgg tcgctgcgct 780
ccctacgccc ccgcgcttcg cgtcggccta tcgcggccgc tggccgctca aaaatggctg 840
gcctacggcc aggcaatcta ccagggcgcg gacaagccgc gccgtcgcca ctcgaccgc 900

```

ggcgcccaca	tcaaggcacc	ctgcctcgcg	cgtttcgggtg	atgacgggtga	aaacctctga	960
cacatgcagc	tcccggagac	ggtcacagct	tgtctgtaag	cggatgccgg	gagcagacaa	1020
gcccgtcagg	gcgcgtcagc	gggtgttggc	gggtgtcggg	gcgcagccat	gaccagtcga	1080
cgtagcgata	gcggagtgtg	tactggctta	actatgcggc	atcagagcag	attgtactga	1140
gagtgcacca	tatgcgggtg	gaaataccgc	acagatgcgt	aaggagaaaa	taccgcatca	1200
ggcgctcttc	cgcttcctcg	ctcactgact	cgctgcgctc	ggtcgttcgg	ctgcggcgag	1260
cggtatcagc	tcactcaaa	gcggtaatac	ggttatccac	agaatcaggg	gataacgcag	1320
gaaagaacat	gtgagcaaaa	ggccagcaaa	aggccaggaa	ccgtaaaaag	gccgcgttgc	1380
tggcggtttt	ccataggctc	cgccccctg	acgagcatca	caaaaaatcg	cgctcaagtc	1440
agaggtggcg	aaaccgcaga	ggactataaa	gataccaggc	gtttccccct	ggaagctccc	1500
tcgtgcgctc	tcctgttccg	accctgccc	ttaccggata	cctgtccgcc	tttctccctt	1560
cgggaaagcgt	ggcgctttct	catagctcac	gctgtaggta	tctcagttcg	gtgtagggtcg	1620
ttcgctccaa	gctgggctgt	gtgcacgaac	cccccgttca	gcccagaccg	tgcgcttat	1680
ccggtaaacta	tcgtcttgag	tcacaaccgg	taagacaacga	cttatcgcca	ctggcagcag	1740
ccactggtaa	caggatttagc	agagcgagg	atgtagcgg	tgctacagag	ttcttgaagt	1800
ggtggcctaa	ctacggctac	actagaagga	cagtatttgg	tatctgcgct	ctgctgaagc	1860
cagttacctt	cggaaaaaga	gttggtagct	cttgcacgg	caaaacaaac	accgctggta	1920
gcgggtgggtt	ttttgtttgc	aagcagcaga	ttacgcgcag	aaaaaaaagg	tctcaagtag	1980
atcctttgat	cttttctacg	gggtctgacg	ctcagtggaa	cgaaaactca	cgtaaaggga	2040
ttttggctcat	gcattctagg	tactaaaaca	attcatccag	taaaaataaa	tatttttattt	2100
tctcccaatc	aggctttgat	cccagtaagt	caaaaaatag	ctcgacatac	tgcttcttccc	2160
cgatatcctc	cctgatcgac	cggacgcaga	aggcaatgtc	ataccacttg	tccgccctgc	2220
cgcttctccc	aagatcaata	aagccactta	ctttgccatc	tttcacaaag	atgttgctgt	2280
ctcccaggtc	gccgtgggaa	aagacaagtt	cctcttcggg	cttttccgtc	tttaaaaaat	2340
catcacagtc	gcgcggatct	ttaaatggag	tgtcttcttc	ccagttttcg	caatccacat	2400
cggccagatc	gttattcagt	aagtaatcca	attcggttaa	gcggctgtct	aagctatttcg	2460
tatagggaca	atccgatatg	tcgatggagt	gaaagagcct	gatgcaactc	gcatacagct	2520
cgataatctt	ttcagggctt	tgttcatctt	catactcttc	cgagcaaaag	acgccatcgg	2580
cctcactcat	gagcagattg	ctccagccat	catgccgttc	aaagtgcagg	acctttggaa	2640
caggcagctt	tccttccagc	catagcatca	tgctcttttc	ccgttccaca	tcataaggtag	2700
tccttttata	ccggctgtcc	gtcattttta	aatatagggt	ttcattttct	cccaccagct	2760
tatatatactt	agcaggagac	attccttccg	tatcttttac	gcagcggtag	ttttcgatca	2820
gttttttcaa	ttccgggtgat	attctcattt	tagccattta	ttatttctct	cctcttttct	2880
acagtattta	aagatacccc	aagaagctaa	ttataacaag	acgaactcca	attcactgtt	2940
ccttgcattc	taaaaacctta	aataccagaa	aacagctttt	tcaaagtgtg	tttcaaagtt	3000
ggcgtataac	atagtatcga	cggagccgat	tttgaaaccg	cggatgatcac	aggcagcaac	3060
gctctgtcat	cgttacaatc	aacatgctac	cctccgcgag	atcatccgtg	tttcaaaacc	3120
ggcagcttag	ttgccgttct	tccgaatagc	atcggtaaca	tgagcaaaag	ctgccgcctt	3180
acaacggctc	tcccgtgac	gccgtcccgg	actgatgggc	tgccctgtatc	gagtgggtgat	3240
tttgtccoga	gctgccggtc	ggggagctgt	gctgtggctg	gtggcaggat	atatgtgtgt	3300
gtaaacaaat	tgacgcttag	acaacttaat	aacacattgc	ggacgttttt	aatgtactga	3360
attaacgccg	aattaatcgg	ggggatctgg	atttttagtac	tggtattttgg	tttttaggaat	3420
tagaaatttt	attgtataga	gtattttaca	aatacaaata	catactaagg	gtttcttata	3480
tgctcaaacac	atgagcgaaa	ccctatagga	accctaattc	ccttatctgg	gaactactca	3540
cacattatta	tggaagaaact	cgagtcaaat	ctcggtgacg	ggcaggaccg	gacggggcgg	3600
taccggcagg	ctgaagtcca	gctgccagaa	acccacgtca	tgccagttcc	cgtgcttgaa	3660
gccggccgcc	cgcagcatgc	cgcggggggc	atatccgagc	gcctcgtgca	tgccgacgct	3720
cgggtcggtg	ggcagcccga	tgacagcgac	cacgctcttg	aagccctgtg	cctccaggga	3780
cttcagcagg	tggtgttaga	gcgtggagcc	cagtcccgtc	cgctgggtggc	gggggggagac	3840
gtacacggtc	gactcggccg	tccagtcgta	ggcgttgctg	gccttccagg	ggcccgcgta	3900
ggcgtatgcc	gcgacctcgc	cgtccacctc	ggcgacgagc	cagggatagc	gctcccgcag	3960
acggagcagg	tcgtccgtcc	actcctgcgg	tcggtagcgg	tcggtagcga	agttgaccgt	4020
gcttgtctcg	atgtagtggg	tgacgatggg	gcagaccgcc	ggcatgtccg	cctcgggtggc	4080
acggcggatg	tcggccgggc	gtcgttctgg	gctcatggta	gactcgagag	agatagattt	4140
gtagagagag	actgggtgatt	tcagcgtgtc	ctctccaaat	gaaatgaact	tccttatata	4200
gaggaaggtc	ttgcgaagga	tagtgggatt	gtgcgtcatc	ccttacgtca	gtggagatat	4260
cacatcaatc	cacttgcctt	gaagacgtgg	ttggaaacgtc	ttctttttcc	acgatgctcc	4320
tcgtgggtgg	gggtccatct	ttgggaccac	tgctggcaga	ggcatcttga	acgatagctc	4380
ttcctttatc	gcaatgatgg	catttgtagg	tgccaccttc	cttttctact	gtccttttga	4440
tgaagtgaca	gatagctggg	caatggaatc	cgaggagggt	tcccgatatt	acctttgtt	4500
gaaaagcttc	aatagccctt	tggtcttctg	agactgtatc	tttgatattc	ttggagtaga	4560
cgagagtgtc	gtgctccacc	atgttatcac	atcaatccac	ttgctttgaa	gacgtgggtg	4620
gaacgtcttc	tttttccacg	atgctcctcg	tgggtggggg	tccatctttg	ggaccactgt	4680
cggcagaggc	atcttgaacg	atagcctttc	ttttatcgca	atgatggcat	ttgtagggtc	4740
caccttccct	ttctactgtc	cttttgatga	agtgacagat	agctgggcaa	tggaaatccga	4800
ggaggtttcc	cgatattacc	ctttgttgaa	aagtctcaat	agccctttgg	tcttctgaga	4860
ctgtatcttt	gatattcttg	gagtagacga	gagtgtcgtg	ctccaccatg	ttggcaagct	4920

gctctagcca atacgcaaac cgcctctccc cgcgcgttgg ccgattcatt aatgcagctg 4980  
gcacgacagg tttcccgaact cagtgcgcgc aacgcaatta atgtgagtta 5040  
gctcactcat taggcacccc aggttttaca ctttatgctt cgggtcgtta tgttgtgtgg 5100  
aattgtgagc ggataacaat ttcacacagg aaacagctat gaccatgatt acgaattcga 5160  
gccttgacta gaggggtcgac ggtatacaga catgataaga tacattgatg agtttggaca 5220  
aaccacaact agaatgcagt gaaaaaaatg ctttatttgg gaaatttgtg atgtatttgc 5280  
tttatttgtta accattataa gctgcaataa acaagttggg gtgggcgaag aactccagca 5340  
tgagatcccc gcgctggagg gtggaatcga aatctcgtag cactgttcag ccgaagccca 5400  
acctttcata gaaggcggcg ttgcccggcg ggtcgcgcag cgggaagtcc cctgtctcct 5460  
cctccgacca ctccggctac atggccggcc cggagggctc ggccaactcc cgcctccacg 5520  
tgtccggcac cacctgggtcc acgaagtccc ggcgcgtgga ggcgaactcc gtggacacga 5580  
caccggcgaa gtcgtcctcc cgcgcgttga ggcacggaaac cgcactggtc gccagggtgt 5640  
cgacggctcc tttcgtctaa gttagtataa aaaagcaggc ttcaatcctg cagggaattcg 5700  
atcgacactc tcgtctactc caagaatata aaagatacag tctcagaaga ccaaagggct 5760  
tttgagactt ggttaatatc ggttaatatc tggatttcca tgggattcca tgccttgcga 5820  
atctgtcact tcatcaaaaag gacagtagaa aaggaaggtg gcacctacaa atgcatcat 6060  
tgcgataaag gaaaggctat cgttcaagat gctctgccc acagtgggtcc caaagatgga 6120  
cccccgacca cgaggagcat cgtggaaaaa gaagacgttc caaccacgtc tccaaggcaa 6180  
gtggattgat gtgataacat ggtggagcac gacactctcg tctactccaa gaatatcaaa 6240  
gatacacttc cagaagacca aagggctatt gagacttttc aacaaagggt aatatcgga 6300  
aacctcctcg cctccattg cccagctatc gttcacttca tcaaaaggac agtagaaaag 6360  
gaaggtggca cctacaaaatg ccatcattgc gataaaggaa aggtatcgt tcaagatgcc 6420  
tctgccgaca gtggtcccaa agatggaccc ccacccacga ggagcatcgt ggaaaaagaa 6480  
gacgttccaa ccacgtcttc aaagcaagtg gatttatgtg atatctccac tgacgtaagg 6540  
gatgacgcac aatccacta tcttctgcaa gacctctctc tatataagga agttcatttc 6600  
atttggagag gacacgctga aatcaccagt ctctctctac aaatctatct ctctcgagct 6660  
ttcgcagatc cgggggggca atgagatatg aaaaagcctg aactcaccgc gactctgtc 6720  
gagaagtctc tgatcgaaaa gttcgcagac gtctccgacc tgatgcagct ctcggagggc 6780  
gaagaatctc gtgctttcag ctctcgatgta ggagggcgtg gatatgtcct gctgggttaa 6840  
agctgcgcgc atgggtttcta tgacattggg gagtttagcg ggcactttgc atcggccgcg 6900  
tccccgattc cagagggtgt caggttgcaa gacctgcctg agagcctgac ctatgtcatc 6960  
tcccgcctg tccgcgaggc tatggatgcg atcgtgcctg aaaccgaact gcccgctgtt 7020  
ctacaacccg cttccggacc gcaaggaatc gttcgaatac ctacatggcg tttttcata 7080  
gggttcggcc ctgatcccca tgtgtatcac tggcaaaactg ccccgagactg ccccgagtc 7140  
tgccgcgatt tgcaggctct tttcggctcc cgagggcgatg ttggttggg atggcagcgc 7200  
acagcggta cgttcggcgg ttcgggctgt atcgccacga cctcgggcgt atatgtccg 7260  
atcttcttct agcttgacgg ggttgacggc atcgggagcc gggactgtcg ggcgtacaca 7320  
aggcatccgg atcttgacgg atcgccacga gtttgcgctc aacaatgtcc tgacggacaa 7380  
gaccaactct atcagagctt caatcgtccg cgatggctgt ggcaaagaa tagagtagat 7440  
cgatgcgacg cgtctggac ctctgctcag agtttctcca taataatgtg tgagtagttc 7500  
agaagcgcgg cctctggac ctctgctcat tttcgtctat gtgttgagca tataagaaac ccttagtatg 7560  
cgccccagca ctctgctcag agtttctcca tttcgtctat tttcgtctat gtttctctca 7620  
gacaagctcg agtatttcca tttcgtctat tttcgtctat gtttctctca 7680  
ccagatcccc cgaattaatt tgactgggaa aacctggcg aatagegaag ggttcagctt 7740  
tacaacgtcg tgactgggaa cagctggcgt gaatggcgaa tggggatcct gagtcacgtt 7800  
cccccttctc tgcgcagcct gccttcagtt tggggatcct agaatcagatt gcatgagcgg 8220  
agaattaagg tggaactgac tgcactaagc acatacgtca aaatatttct agagtctcat 8280  
tgagctagc atcagctagc gtatccaatt atcgaattcc gagttgtccc tgggagagg 8340  
gtggagagg gtggaaaaac gcttttcaag agggatcgt ctgaagtcaa gtttgaggga 8400  
atccacagat gcaggagagg gtttgaggga gacacccctc tcaacaggat 8460  
cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc 8520  
cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc 8580  
cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc 8640  
cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc 8700  
cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc 8760  
cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc 8820  
cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc 8880  
cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc cagctgcctc 8940

-56-

tcaaggaggga	cggaacacatc	ctcggccaca	agttggaata	caactacaac	tcccacaacg	9000
tatacatcat	ggccgcacaag	caaaagaacg	gcatcaaagc	caacttcaag	acccgccaca	9060
acatcgaaga	cggcgcggtg	caactcgctg	atcattatca	acaaaataact	ccaattggcg	9120
atggccctgt	cctttttacca	gacaaccatt	acctgtccac	acaatctgcc	ctttcgaaag	9180
atcccaacga	aaagagagac	cacatggtcc	ttcttgagtt	tgtaacagct	gctgggatta	9240
cacatggcat	ggatgaacta	tacaaagcta	gccaccacca	ccaccaccac	gtgtgaattg	9300
gtgaccagct	cgaatttccc	cgatcgttca	aacatttggc	aataaaagttt	cttaagattg	9360
aatcctgttg	cgggtcttgc	gatgattatc	atataatttc	tggtgaatta	cgtaaagcat	9420
gtataaatta	acatgtaatg	catgacgtta	tttatgagat	gggtttttat	gatttagagtc	9480
ccgcaattat	acatttaata	cgcgatagaa	aacaaaatat	agcgcgcaaa	ctaggataaa	9540
ttatcgcgcg	cgggtgtcatc	tatgttacta	gatcgggaaat	taaaactatca	gtgtttgaca	9600
ggatatattg	gcgggttaaac	ctaagagaaa	agagcgttta	ttagaataaac	ggatatttaa	9660
aagggcggtga	aaagggtttat	ccgttcgtcc	atttgtatgt	gcatgccaac	cacagggttc	9720
ccctcgggat	caaagtaactt	tgatccaaac	cctccgctgc	tatagtgcag	tccgcttctg	9780
acgttcagtg	cagccgtctt	ctgaaaacga	catgtcgcac	aagtcctaag	ttacgcgaca	9840
ggctgccgcc	ctgccctttt	cctggcgttt	tcttgtcgcg	tgtttttagtc	gcataaagta	9900
gaataacttg	gactagaacc	ggagacatta	cgccatgaac	aagagcgccg	ccgctggcct	9960
gctgggctat	gcccgcgtca	gcaccgacga	ccaggacttg	accaaccaac	gggcccgaact	10020
gcacgcggcc	ggctgcacca	agctgttttc	cgagaagatc	accggcacca	ggcgcgaccg	10080
cccggagctg	gccaggatgc	ttgaccacct	acgcccgggc	gacgttgtga	cagtggaccag	10140
gctagaccgc	ctggcccgca	gcaccgcga	cctactggac	attgccgagc	gcataccagga	10200
ggccggcgcg	ggcctgcgta	gcctggcaga	gccgtgggccc	gacaccacca	cgccggcgccg	10260
ccgcatgggtg	ttgaccgtgt	tcgcccgcac	tgccgagttc	gagcggtccc	taatcatcga	10320
ccgcaccccg	agcgggcgcg	agggccgcaa	ggcccagggc	gtgaagtgtt	gcccccgccc	10380
taccctcacc	ccggcacaga	tcgcgacgc	ccgcgagctg	atcgaccagg	aaggccgcac	10440
cgtgaaagag	gcccgtgcac	tgcttggcgt	gcactcgctc	accctgtacc	gcgcacttga	10500
gcgacgcgag	gaagtgcgc	ccaccgaggg	caggcgccgc	ggtgccttcc	gtgaggagca	10560
attgaccgag	gcgcagcccc	tgccggccgc	cgagaatgaa	cgccaagagg	aacaagcatg	10620
aaaccgcacc	aggacggcca	ggacgaaccg	tttttcatta	ccgaagagat	cgaggcgagat	10680
atgatcgcg	ccgggtacgt	gttcgagccc	ccgcgcacg	tctcaaccgt	gcggctgcgt	10740
tgaaggttat	ccggtttgtc	tgatgccaa	ctggcgccct	ggccggccag	cttgcccgct	10800
gaagaaaccg	agcgcgcgcg	tctaaaaagg	tgatgtgtat	ttgagtaaaa	cagcttgcgt	10860
catgcggtcg	ctgcgtatat	gatgcgatga	gtaaataaac	aaatacgcaa	ggggaacgca	10920
tgaaggttat	cgctgtactt	aaccagaaa	gcgggtcagg	caagacgacc	atcgcaaccc	10980
atctagcccc	cgccctgcaa	ctcgccgggg	ccgatgttct	gttagtgcgt	tccgatcccc	11040
agggcagtg	ccgcgattgg	gcggccgtgc	gggaagatca	accgctaacc	gttgtcgga	11100
tcgacggccc	gacacgccc	gcgcagctga	agggcatcgg	ccggcgcgac	ctgtagtga	11160
tcgacgggag	gccccaggcg	gcggacttgg	ctgtgtccgc	gatcaaggca	gcccacttgc	11220
tgctgatccc	ggtgcagcca	agcccttacg	acatatgggc	caccgcccag	ctggtggagc	11280
tggttaagca	gcgcattgag	gtcacggatg	gaaggctaca	agcgcccttt	gtcgtgtcgc	11340
gggcgatcaa	aggcacgcgc	atcgccgggtg	aggttgccga	ggcgctggcc	gggtacgagc	11400
tgcccattct	tgagtcccgt	atcacgcagc	gcgtgagcta	cccaggcaact	gcccgcgcgcg	11460
gcacaaccgt	tcttgaatca	gaaccggagg	gcgacgctgc	ccgcgaggtc	caggcgctgg	11520
ccgctgaaat	taaatcaaaa	ctcatttgag	ttaatgaggt	aaagagaaaa	tgagcaaaaag	11580
cacaaacacg	ctaagtgcgc	gccgtccgag	cgcacgcagc	agcaaggctg	caacgttggc	11640
cagcctggca	gacacgccc	ccatgaagcg	ggccaacttt	cagttgccc	cggaggatca	11700
caccaagctg	aagatgtacg	cggtacgcca	aggcaagacc	attaccgagc	tgctatctga	11760
atacatcgcg	cagctaccag	agtaaatgag	caaatgaata	aatgagtaga	tgaatttttag	11820
cggctaaagg	aggcggcatg	gaaaatcaag	aacaaccagg	caccgacgcc	gtggaatgcc	11880
ccatgtgtgg	aggaaacgggc	ggttggccag	gcgtaagcgg	ctgggttgtc	tgccggccct	11940
gcaatggcac	tggaaacccc	aagcccagg	aatcgccgtg	acggtcgcaa	accatccggc	12000
ccggtacaaa	tcggcgcggc	gctgggtgat	gacctggtgg	agaagttgaa	ggccgcgcag	12060
gcccggcagc	ggcaacgcac	cgaggcagaa	gcacgccccg	gtgaatcgtg	gcaagcggcc	12120
gctgatcgaa	tccgcaaa	atcccggcaa	ccgcggcgag	ccggtgcgcc	gtcgattagg	12180
aagccgccc	agggcgacga	gcaaccagat	tttttcgttc	cgatgctcta	tgacgtgggc	12240
acccgcgata	gtcgcagcat	catggacgtg	gccgttttcc	gtctgtcgaa	gcgtgaccga	12300
cgagctggcg	aggtgatccg	ctacgagctt	ccagacgggc	acgtagaggt	ttccgcaggg	12360
ccggccggca	tggccagtg	gtgggattac	gacctggtac	tgatggcggt	ttcccatcta	12420
accgaatcca	tgaaccgata	ccgggaagg	aagggagaca	agcccggccg	cgtgttccgt	12480
ccacacgttg	cggacgtact	caagttctgc	cggcgagccg	atggcggaag	gcagaaagac	12540
gacctggtag	aaacctgcat	tcggttaaac	accacgcacg	ttgccatgca	gc	12592

&lt;210&gt; 96

&lt;211&gt; 3357

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence



-57-

&lt;220&gt;

&lt;223&gt; pGEMEasyNOS Plasmid

&lt;400&gt; 96

tatcactagt	gaatttcg	ccgctgcag	gtcgaccata	tgggagagct	cccaacgcgt	60
tggatgcata	gcttgagt	tctatagt	cacctaaata	gcttggcgta	atcatgggtca	120
tagctgtttc	ctgtgtgaaa	ttgttatccg	ctcacaaatc	cacacaacat	acgagccgga	180
agcataaagt	gtaaagccctg	gggtgcctaa	tgagttagct	aactcacatt	aattgcgttg	240
cgctcactgc	ccgctttcca	gtcgggaaac	ctgtcgtgcc	agctgcatta	atgaatcggc	300
caacgcgcgg	ggagaggcgg	tttgcgtatt	gggcgtctct	ccgcttcctc	gctcactgac	360
tcgctgcgct	cggtcgttcg	gctgcggcga	gcggtatcag	ctcactcaaa	ggcggtaata	420
cggttatcca	cagaatcagg	ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	480
aaggccagga	accgtaaaaa	ggccgcgttg	ctggcgtttt	tccataggct	ccgccccct	540
gacgagcatc	acaaaaatcg	acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	600
agataccagg	cgtttccccc	tggaaagctcc	ctcgtgcgct	ctcctgttcc	gaccctgccg	660
cttaccggat	acctgtccgc	ctttctccct	tcgggaagcg	tggcgctttc	tcatagctca	720
cgctgtagg	atctcagttc	gggtgtaggtc	gttcgtccca	agctgggctg	tgtgcacgaa	780
cccccgcttc	agcccgaccg	ctgcgcctta	tcgggtaaact	atcgtcttga	gtccaaacccg	840
gtaagacacg	acttatcgcc	actggcagca	gccactggta	acaggattag	cagagcgagg	900
tatgtaggcg	gtgctacaga	gttcttgaag	tgggtggccta	actacggcta	cactagaaga	960
acagtatttg	gtaactgcgc	tctgctgaag	ccagttacct	tcggaaaaaag	agttggtagc	1020
tcttgatccg	gcaaacaaaac	caccgctggt	agcgggtggt	tttttgtttg	caagcagcag	1080
attacgcgca	gaaaaaaagg	atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	1140
gctcagtgga	acgaaaaactc	acgttaaagg	attttgggtca	tgagattatc	aaaaaggatc	1200
ttcacctaga	tcctttttaa	ttaaaaatga	agttttaaat	caatctaaag	tatatatgag	1260
taaaacttgg	ctgacagtta	ccaatgctta	atcagtgagg	cacctatctc	agcgatctgt	1320
ctatttcggt	catccatagt	tgctgactc	ccgctcgtgt	agataaactac	gatacgggag	1380
ggcttaccat	tgcccccag	tgctgcaatg	ataccgcgag	accacgcctc	accggctcca	1440
gattttatcag	caataaacca	gccagccgga	agggccgagc	gcagaagtgg	tcctgcaact	1500
ttatccgcct	ccatccagtc	tattaattgt	tgccgggaag	ctagagtaag	ctagtcgcca	1560
gttaaatagt	tgcccaacgt	tggtgccatt	gctacaggca	tcgtgggtgtc	acgctcgctc	1620
tttggtatgg	cttcattcag	ctccggttcc	caacgatcaa	ggcgagttac	atgatcccc	1680
atgttgtgca	aaaaagcgg	tagctccttc	ggctctccga	tcgttgtcag	aagtaagttg	1740
gccgcagtg	tatcactcat	ggttatggca	gtcactcata	attctcttac	tgtcatggca	1800
tccgtaagat	gcttttctgt	gactggtgag	tactcaacca	agtcattctg	agaatagtgt	1860
atgcggcgac	cgagttgtct	ttgcccggcg	tcaataccgg	ataataccgc	gccacatagc	1920
gtaactttta	aagtgtcat	cattggaaaa	cgttcttcgg	ggcgaaaact	ctcaaggatc	1980
ttaccgctgt	tgagatccag	ttcgatgtaa	cccactcgtg	cacccaactg	atcttcagca	2040
tcttttactt	tcaccagcgt	ttctgggtga	gcaaaaaacag	gaaggcaaaa	tgccgcaaaa	2100
aagggaataa	gggcgacacg	gaaatgttga	atactatac	tcttcctttt	tcaatattat	2160
tgaagcattt	atcagggtta	ttgtctcatg	agcggatata	tatttgaatg	tatttagaaa	2220
aataaacaaa	taggggttcc	gcgcacattt	ccccgaaaag	tgccacctga	tgccggtgtga	2280
aataaccgac	agatgcgtaa	ggagaaaaata	ccgcatacagg	aaattgttaag	cgttaatatt	2340
ttgttaaaat	tcgcgttaaa	tttttgttaa	atcagctcat	tttttaacca	ataggccgaa	2400
atcggaaaaa	tcctttataa	atcaaaagaa	tagaccgaga	taggggttgag	tggtgttcca	2460
gtttggaaca	agagtccact	attaaagaac	gtggactcca	acgtcaaagg	gcgaaaaaac	2520
gtctatcagg	gcgatggccc	actacgtgaa	ccatcacctc	aatcaagttt	tttggggctg	2580
aggtgccgta	aagcactaaa	tcggaaccct	aaagggagcc	cccgatttag	agcttgacgg	2640
ggaaaagccg	cgaacgtggc	gagaaaggaa	gggaagaaag	cgaaaggagc	ggcgctagg	2700
gcgctggcaa	gtgtagcgg	cacgctgcgc	gtaaccacca	cacccgcgcg	gcttaatgcg	2760
ccgctacagg	gcgcgtccat	tcgccattca	ggctgcgcaa	ctgttgggaa	ggcgcatcgg	2820
tcggggcctc	ttcgctatta	cgccagctgg	cgaaaggggg	atgtgctgca	aggcgattaa	2880
gttgggtaac	gccaggggtt	tcccagtcac	gacgttgtaa	aacgacggcc	agtgaattgt	2940
aatacgactc	actatagggc	gaattggg	cgacgtcgca	tgctcccggc	cgccatggcg	3000
gccgcgggaa	ttcgattctc	gagatccggt	gcagattatt	tggattgaga	gtgaatatga	3060
gactctaatt	ggataccgag	gggaatttat	ggaacgtcag	tggagcattt	ttgacaagaa	3120
atatttgcta	gctgatagtg	accttaggcg	acttttgaac	gcgcaataat	ggtttctgac	3180
gtatgtgctt	actcatttaa	actccagaaa	cccgcggctg	agtggctcct	tcaacggttg	3240
ggttctgtca	gttccaaacg	taaaacggct	tgtcccgcgt	catcgccggg	ggtcataaac	3300
tgactccctt	aattctccgc	tcatgatcag	attgtcgttt	ccgccttca	gtctaga	3357

&lt;210&gt; 97

&lt;211&gt; 10122

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; p1302NOS Plasmid

&lt;400&gt; 97

catggttagat	ctgactagta	aaggagaaga	acttttctact	ggagttgtcc	caatttcttgt	60
tgaattagat	ggtgatgtta	atgggcacaa	attttctgtc	agtggagagg	gtgaagggtga	120
tgcaacatac	ggaaaactta	cccttaaatt	tatttgcact	actggaaaac	tacctgttcc	180
gtggccaaca	cttgtcacta	ctttctctta	tgggtgttcaa	tgctttttcaa	gatacccca	240
tcatatgaag	cggcacgact	tcttcaagag	cgccatgcct	gagggatacg	tgaggagag	300
gaccatcttc	ttcaaggacg	acgggaacta	caagacacgt	gctgaagtca	agtttgaggg	360
agacaccctc	gtcaacagga	tcgagcttaa	gggaatcgat	ttcaaggagg	acggaaacat	420
cctcgccac	aagttggaat	acaactacaa	ctcccacaac	gtatacatca	tggccgacaa	480
gcaaaagaac	ggcatcaaag	ccaacttcaa	gacccgccac	aacatcgaag	acggcgcgct	540
gcaactcgct	gatcattatc	aacaaaatc	tccaattggc	gatggccctg	tccttttacc	600
agacaaccat	tacctgtcca	cacaatctgc	cctttcgaaa	gatcccaacg	aaaagagaga	660
ccacatggtc	cttcttgagt	ttgtaacagc	tgctgggatt	acacatggca	tgatgaact	720
atacaaagct	agccaccacc	accaccacca	cgtgtgaatt	ggtgaccagc	togaatttcc	780
ccgatcgctc	aaacatttgg	caataaagtt	tcttaagatt	gaatcctgtt	gccggtcttg	840
cgatgattat	acataaattt	ctgttgaatt	acgttaagca	tgtaataatt	aacattgta	900
gcatgacggt	atttatgaga	tgggttttta	tgattagagt	cccgaatta	tacatttta	960
acgcataga	aaacaaaata	tagcgcgcaa	actaggataa	attatcgcg	gcggtgtcat	1020
ctatgttact	agatcgggaa	ttaaactatc	agtgtttgac	aggatatatt	ggcggtgtaa	1080
cctaagagaa	aagagcggtt	attagaataa	cggatattta	aaagggcggt	aaaagggtta	1140
tccgttcgtc	catttgtatg	tgcatgccaa	ccacagggtt	cccccgggg	tcaaagtact	1200
ttgatccaac	ccctcgcgtg	ctatagtcca	gtcggcttct	gacgttcagt	gcagccgct	1260
tctgaaaacg	acatgtcgca	caagtcctaa	gttacgcgac	aggctgccc	cctgcccctt	1320
tccgtggcgt	ttcttgcgc	gtgttttagt	cgcataaagt	agaatacttg	cgactagaac	1380
cgagacatt	acgcctgaa	caagagcgcc	gcccgtggcc	tgctgggcta	tgcccgcgtc	1440
agcaccgacg	accaggactt	gaccaaccaa	agggccgaac	tgacgcggc	cggctgcacc	1500
aagctgtttt	ccgagaagat	caccggcacc	agggcgacc	gcccggagct	ggccaggatg	1560
cttgaccacc	tacgccttgg	cgactgtgtg	acagtgaacca	ggctagaccg	cctggccgcg	1620
agcaccgcg	acattcgga	cattgcgag	cgcattccagg	agggcgggc	gggctgctg	1680
agcctggcag	agcctggggc	cgacaccacc	acggcgggcg	gcccgcaggt	ggtgaccgtg	1740
ttcgccggca	ttgcccaggt	cgagcggttc	ctaatctcg	accgcacccg	gagcgggcgc	1800
gaggccggca	aggcccaggg	cgtgaaggtt	ggcccccgcc	ctaccctcac	ccgggcacag	1860
atcgccgacg	cccgcgagct	gatcgaccag	gaaggccgca	ccgtgaaaga	ggcggttgca	1920
ctgcttggcg	tgcattcgct	gaccctgtac	cgcgcatctg	agcgcagcga	ggaagtgcag	1980
cccaccgcgt	ccaggcgggc	cggtgccttc	cgtgaggacg	cattgaccga	ggccgacgcc	2040
ctggcgggcg	ccgagaatga	acgccaagag	gaacaagcat	gaaaccgcac	caggacggcc	2100
aggacgaacc	gtttttcatt	accgaagaga	tcgaggcgga	gatgatcgcg	gcccgggtac	2160
tggttcgagc	gcccgcgcac	gtctcaaccg	tgcggctgca	tgaaatcctg	gcccggttgt	2220
ctgatgccaa	gctggcgggc	tggccggcca	gcttggccgc	tgaagaaacc	gagcgccgcc	2280
gtctaaaaag	gtgatgtgta	tttgagtaaa	acagcttgcg	tcatgcggtc	gctgcgtata	2340
tgatcgatg	agtaaatgaa	caaatacgca	aggggaacgc	atgaagggtt	tcgctgtact	2400
taaccagaaa	ggcgggtcag	gcaagacgac	catcgcaacc	catctagccc	gccccttgca	2460
actcgccggg	gcccgatgtc	tgttagtcca	ttccgatccc	cagggcagtg	cccgcgattg	2520
ggcgcccggt	cggaagatc	aaccgctaac	cgttgtcggc	atcgaccgcc	cgacgattga	2580
ccgcgacgtg	aaggccatcg	gcccgcgcga	cttcgttagt	atcgacggag	cgcccaggcc	2640
ggcggaactg	gctgtgtccg	cgatcaaggc	agccgacttc	gtgctgattc	cggtgcagcc	2700
aagcccttac	gacatatggg	ccaccgcgca	cctgggtggag	ctgggttaagc	agcgatttga	2760
ggtcacggat	ggaaggctac	aagcggcctt	tgtcgtgtcg	cgggcgatca	aaggcacgcg	2820
catcgccggt	gaggttgccg	agggcgctgg	cgggtacgag	ctgcccattc	ttgagtcctg	2880
tatcacgcag	cgcgtgagct	acccaggcac	tgccgcggcc	ggcacaaccg	ttcttgaatc	2940
agaaccgcag	ggcgacgctg	cccgcgaggt	ccaggcgctg	gcccgtgaaa	ttaaatcaaa	3000
actcatttga	gttaatgagg	taaagagaaa	atgagcaaaa	gcacaaacac	gctaagtgcc	3060
ggccgtccga	gcgcacgcag	cagcaaggct	gcaacgttgg	ccagcctggc	agacacgcca	3120
gccatgaagc	gggtcaactt	tcagttgccc	gcgaggatc	acaccaagct	gaagatgtac	3180
gcggtacgcc	aaggcaagac	cattaccgag	ctgctatctg	aatacatcgc	gcagctacca	3240
gagtaaatga	gcaaatgaat	aaatgagtga	atgaatttta	gcccgtaaaag	gaggcggtat	3300
ggaaaatcaa	gaacaaccag	gcaccgacgc	cgtggaatgc	cccatgtgtg	gaggaacggg	3360
cggttggcca	ggcgtaagcg	gctgggttgt	ctgcgggccc	tgcaatggca	ctggaacccc	3420
caagccgcag	gaactcggcg	gacggtcgca	aaccatccgg	cccggtaaaa	atcgccgccc	3480
cgctgggtga	tgacctgggtg	gagaagttga	agggcgcgca	ggccgcccag	cggcaacgca	3540
tcgaggcaga	agcagcggcc	ggtgaatcgt	ggcaagcggc	cgctgatcga	atccgcaaa	3600
aatcccgcga	acccgcggca	gccggtgcgc	cgcggttag	gaagccgccc	aaggccgacg	3660
agcaaccaga	ttttttcggt	ccgatgctct	atgacgtggg	caccgcgcat	agtcgcagca	3720
tcatggacgt	ggccggtttt	cgtctgtcga	agcgtgaccg	acgagctggc	gaggtgatcc	3780
gctacgagct	tccagacggg	cacgtagagg	tttccgcagg	gcccggccgg	atggccagtg	3840



tgtgggatta	cgacctggta	ctgatggcgg	tttcccatct	aaccgaatcc	atgaaccgat	3900
accgggaagg	gaagggagac	aagcccggcc	gcgtgttccg	tccacacgtt	gcggacgtac	3960
tcaagttctg	ccggcgagcc	gatggcggaa	agcagaaaaga	cgacctggta	gaaacctgca	4020
ttcggttaaa	caccacgcac	gttgccatgc	agcgtacgaa	gaaggccaag	aacggccgcc	4080
tggtagcggg	atccgagggg	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	4140
ccgggcccgc	ggagtagatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	4200
aaggcaagaa	cccggacgtg	ctgacgggtt	accccgatta	ctttttgatc	gatcccggca	4260
tcggccggtt	tctctaccgc	ctggcacgcc	gcgcgcgagg	caaggcagaa	gccagatggg	4320
tgttcaagac	gctctacgaa	cgagtgggca	gcgcgcgaga	gttcaagaag	ttctgtttca	4380
ccgtgcgcaa	gctgatcggg	tcaaatgacc	tgccggagta	cgatttgaag	gaggaggcgg	4440
ggcaggtctg	cccgatccta	gtcatgcgct	accgcaacct	gatcgagggc	gaagcatccg	4500
ccggttccta	atgtacggag	cagatgctag	ggcaaatgtg	cctagcaggg	gaaaaagggt	4560
gaaaagggtc	ctttcctgtg	gatagcacgt	acattgggaa	cccaaagccg	tacattggga	4620
accggaaccc	gtacattggg	aacccaaagc	cgtacattgg	gaaccgggtc	cacatgtaag	4680
tgactgatat	aaaagagaaa	aaaggcgatt	tttcgccta	aaactcttta	aaacttatta	4740
aaactcttaa	aaccgcctcg	gcctgtgcat	aactgtctgg	ccagcgcaca	gccgaagagc	4800
tgcaaaaagc	gcctaccctt	cggtcgctgc	gctccctacg	ccccgccgct	tcgctcgggc	4860
ctatcgccgc	cgctggccgc	tcaaaaatgg	ctggcctacg	gccaggcaat	ctaccagggc	4920
gcggacaagc	cgcgcgctcg	ccactcgacc	gcgcgcgcc	acatcaaggc	accctgcctc	4980
gcgcgtttcg	gtgatgacgg	tgaaaacctc	tgacacatgc	agctcccggg	gacgggtcac	5040
gcttgtctgt	aagcggatgc	cgggagcaga	caagcccgct	agggcgcgct	agcgggtggt	5100
ggcgggtgtc	ggggcgagc	catgacccag	tcacgtagcg	atagcggagt	gtatactggc	5160
ttaactatgc	ggcatcagag	cagattgtac	tgagagtgc	ccatagcg	tgtgaaatac	5220
cgcacagatg	cgtaaggaga	aaataccgca	tcaggcgctc	ttccgcttcc	tcgctcactg	5280
actcgctgct	ctcggtcggt	cggtgcggc	gagcggatc	agctcactca	aaggcggtaa	5340
tacggttatc	cacagaatca	ggggataacg	caggaaaagaa	catgtgagca	aaaggccacc	5400
aaaaggccag	gaaccgtaaa	aaggcccgct	tttcgctggt	tttccatagg	ctccgcccc	5460
ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagagggt	gcgaaacccg	acaggactat	5520
aaagatacca	ggcggtttccc	cctggaagct	ccctcggtgc	ctctcctggt	ccgaccctgc	5580
cgcctaccgg	atacctgtcc	gcctttctcc	cttcgggaag	cggtggcgctt	ttctcatagt	5640
cacgctgtag	gtatctcagt	tcgggtgtagg	tcgttcgctc	caagctgggc	tgtgtgcacg	5700
aaccccccg	tcagcccgac	cgctgcgcct	tatccggtaa	ctatcgctctt	gagtccaacc	5760
cggttaagaca	cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	5820
ggtaggtagg	gagttcttga	gagttcttga	agtggtggcc	taactacggc	tacactagaa	5880
ggacagtat	tggtatctgc	gctctgctga	agccagttac	cttcggaaaa	agagttggta	5940
gctcttgatc	cggaacacaa	accaccgctg	gtagcgggtg	tttttttgtt	tgcaagcagc	6000
agattacgcg	cagaaaaaaa	ggatctcaag	aagatccctt	gatcttttct	acgggggtctg	6060
acgctcagtg	gaacgaaaaa	tcacgttaag	ggatttttgt	catgcattct	aggtactaaa	6120
acaattcatc	cagtaaaaata	taataatttta	tttttccca	atcaggcttg	atccccagta	6180
agtcacaaaa	tagctcgaca	tactgttctt	ccccgatatc	ctccctgatc	gaccggagcg	6240
agaaggcaat	gtcataccac	ttgtccgccc	tgccgcttct	cccaagatca	ataaagccac	6300
ttactttgcc	atcttttcaca	aagatgttgc	tgctctccag	gtcgcgctgg	gaaaagacaa	6360
gttctctctc	gggcttttcc	gtctttaaaa	aatcatacag	ctcgcgcgga	tctttaaatg	6420
gagtgctctc	ttcccagttt	tcgcaatcca	catcggccag	atcggttatc	agtaagtaat	6480
ccaattcggc	taagcggctg	tctaagctat	tcgtataggg	acaatccgat	atgtcgatgg	6540
agtgaagag	cctgatgcac	tcgcgataca	gctcgataat	cttttcaggg	ttcggtcatt	6600
cttcatactc	ttccgagcaa	aggacgcat	cggcctcact	catgagcaga	ttgctccagc	6660
catcatgcgg	ttcaaagtgc	aggaccttgc	gaacaggcag	ctttccttcc	agccatagca	6720
tcattgtcct	ttcccgttcc	acatcatagg	tggtcccttt	ataccggctg	tccgtcattt	6780
ttaaatatag	gttttcattt	tctcccacca	gcttatatac	cttagcagga	gacattcctt	6840
ccgtatcttt	tacgcagcgg	tatttttcga	tcagtttttt	caattccggg	gatattctca	6900
tttttagccat	ttattatttc	cttccctctt	tctacagtat	ttaaagatac	cccaagaagc	6960
taattataac	aagacgaact	ccaattcact	gttccctgca	ttctaaaacc	ttaaatacca	7020
gaaaacagct	ttttcaaagt	tgttttcaaa	gttggtgat	aacatagtat	cgacggagcc	7080
gattttgaaa	ccgcgggtgat	cacaggcagc	aacgctctgt	catcgttaca	atcaacatgc	7140
tacctccgc	gagatcatcc	gtgtttcaaa	cccggcagct	tagttgccgt	tcttccgaat	7200
agcatcggtg	acatgagcaa	agtcgtccgc	cttataacgg	ctctcccgtc	gacgccgtcc	7260
cggactgatg	ggctgctgtg	atcgagtggt	gtattttgtc	cgagctgccc	gtcggggagc	7320
tggttggtgg	ctggtggcag	gatatattgt	ggtgtaaaaa	aattgacgct	tagacaactt	7380
aataacacat	tgccgagcgt	tttaattgtac	tgaattaacg	ccgaattaat	tcggggggatc	7440
tggatttttag	tactggattt	tggttttagg	tttattgata	gaagtatttt	gaagtatttt	7500
acaaatacaa	atacatacta	agggtttctt	atatgctcaa	cacatgagcg	aaaccctata	7560
ggaacccata	ttcccttacc	tgggaaactac	tcacacatta	ttatggagaa	actcgagctt	7620
tcgcatcgac	agatccggtc	ggcatctact	ctatttcttt	gccctcggac	gagtgctggg	7680
gcgtcggttt	ccactatcgg	cgagtacttc	tacacagcca	tcggtccaga	cggccgcgct	7740
tctgcggggc	atgtgtgtac	gcccagacgt	cccggctccg	gatcggacga	ttgcgtcgca	7800
tcgacctgc	gcccagctgc	catcatcgaa	attgcgctca	accaagctct	gatagagttg	7860

-60-

```

gtcaagacca atgaggagca tatacgcccc gagtcgtggc gatcctgcaa gctccggatg 7920
cctccgctcg aagtagcgcg tctgctgctc catacaagcc aaccacggcc tccagaagaa 7980
gatgtttggcg acctcgattt gggaatcccc gaacatcgcc tcgctccagt caatgaccgc 8040
tgttatgctgg ccattgtccg tcaggacatt gttggagccg aaatccgctg gcacgaggtg 8100
ccggacttcg gggcagtcct cggcccaag catcagctca tcgagagcct gcgcgacgga 8160
cgactgacg gtgtcgcca tcacagtttg ccagtgtatc acatggggat cagcaatcgc 8220
gcatatgaaa tcacgccatg tagtgtattg accgattcct tgcggtccga atgggcccga 8280
cccgctcgctc tggctaagat cggccgcagc gatcgcatcc atagcctccg cgaccggttg 8340
tagaacagcg ggcagttcgg tttcaggcag gtcttgcaac gtgacaccct gtgcacgggtg 8400
ggagatgcaa taggtcaggc tctcgctaaa ctccccaatg tcaagcactt ccggaatcgg 8460
gagcgcggcc gatgcaaagt gccgataaac ataacgatct ttgtagaaac catcggcgca 8520
gctattttacc cgcaggacat atccacgccc tcctacatcg aagctgaaag cagagattc 8580
ttcgccctcc gagagctgca tcaggtcgga gacgctgtcg aacttttcga tcagaaactt 8640
ctcgacagac gtccggtgga gttcaggctt tttcatatct cattgcccc ccggatctgc 8700
gaaagctcga gagagataga tttgtagaga gagactgggtg atttcagcgt gtcctctcca 8760
aatgaaatga acttccttat atagaggaag gtcttgcgaa ggatagtggg attgtgcgtc 8820
atcccttacg tcagtggaga tatcacatca atccacttgc tttgaagacg tgggtggaac 8880
gtcttctttt tccacgatgc tctcgctggg tgggggtcca tctttgggac cactgtcggc 8940
agaggcatct tgaacgatag cctttccttt atcgcaatga tggcatttgt aggtgccacc 9000
ttccttttct actgtccttt tgatgaagtg acagatagct gggcaatgga atccgaggag 9060
gtttcccgat attacccttt gttgaaaagt ctcaatagcc ctttgggtct ctgagactgt 9120
atctttgata ttcttggagt agacgagagt gtcgtgctcc accatgttat cacatcaatc 9180
cacttgcttt gaagacgtgg ttggaacgtc tcttttttcc acgatgctcc tctgtgggtg 9240
gggtccatct ttgggaccac tgtcggcaga ggcacttga acgatagcct ttcctttatc 9300
gcaatgatgg catttgtagg tgccaccttc gtctctttga tgaagtgaca 9360
gatagctggg caatggaate cgaggaggtt tcccgatatt accctttgtt gaaaagtctc 9420
aatagccctt tggcttctg agactgtatc tttgatattc ttggagtga cgagagtgtc 9480
gtgctccacc atgttggcaa gctgctctag ccaatacgca aaccgcctct ccccgcgctg 9540
tgcccgattc attaatgcag ctggcagcac aggtttcccg actggaaagc gggcagtgag 9600
cgcaacgcaa ttaatgtgag ttagctcact cattagcac cccaggcttt acactttatg 9660
cttccggctc gtatgttgtg tgggaattgtg agcggataac aatttcacac aggaaacagc 9720
tatgaccatg attacgaatt cgagctcggg acccggggat cctctagact gaaggcggga 9780
aacgacaatc tgatcatgag cggagaatta gggagctcac gttatgacc ccccgcatga 9840
cgcgggacaa gccgttttac gtttggaaact gacagaaccg caacgttgaa ggagccactc 9900
agccgcgggt ttctggagtt taatgagcta agcacatagc tcagaaacca ttattgcgcg 9960
ttcaaaagtc gcctaaggtc actatcagct taccataatt tcttgtcaaa aatgtctcac 10020
tgacgttcca taaattccc tcggtatcca attagagtct catattcact ctcaatccaa 10080
ataatctgca ccggatctcg agaatcgaat tccgcggcc gc 10122

```

<210> 98  
 <211> 621  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> N. tabacum rDNA intergenic spacer (IGS) sequence

<300>  
 <308> Genbank #Y08422  
 <309> 1997-10-31

```

<400> 98
gtgctagcca atgtttaaca agatgtcaag cacaatgaat gttgggtggtt ggtgggtcgtg 60
gctggcggtg gtggaaaatt gcggtggttc gagcggtagt gatcggcgat ggttgggtgtt 120
tgacgcggtg tttgatatacg gaatcactta tgggtggttgt cacaatggag gtgcgtcatg 180
gttattgggtg gttgggtcatc tatatatttt tataataata ttaagtattt tacctatttt 240
ttacataattt tttattaaat ttatgcattg tttgtatttt taaatagttt ttatcgctact 300
tgttttataaa aatatttttat tatttttatgt gtttatattat tacttgatgt attggaaatt 360
ttctccattg ttttttctat attttataata attttcttat ttttttttgt tttattatgt 420
attttttcgt tttataataa atattttatta aaaaaaatat tatttttgta aaatatatca 480
tttacaatgt tttaaaagtc tttgtgaata tattagctaa gttgtacttc tttttgtgca 540
tttgggtgtg tacatgtcta ttatgattct ctggccaaaa catgtctact cctgtcactt 600
gggttttttt ttttaagaca t 621

```

<210> 99  
 <211> 25  
 <212> DNA

-61-

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; NTIGS-F1 Primer

&lt;400&gt; 99

gtgctagcca atgtttaaca agatg

25

&lt;210&gt; 100

&lt;211&gt; 28

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; NTIGS-R1 Primer

&lt;400&gt; 100

atgtcttaaa aaaaaaacc caagtgc

28

&lt;210&gt; 101

&lt;211&gt; 233

&lt;212&gt; DNA

&lt;213&gt; Mus Musculus

&lt;300&gt;

&lt;308&gt; Genbank #V00846

&lt;309&gt; 1989-07-06

&lt;400&gt; 101

gacctggaat	atggcgagaa	aactgaaaat	cacggaaaat	gagaaataca	cacttttagga	60
cgtagaatat	ggcgaggaaa	actgaaaaag	gtggaaaatt	tagaaatgtc	cactgttagga	120
cgtagaatat	ggcaagaaaa	ctgaaaatca	tggaaaatga	gaaacatcca	cttgacgact	180
tgaaaaatga	cgaaatcact	aaaaaacgtg	aaaaatgaga	aatgcacact	gaa	233

&lt;210&gt; 102

&lt;211&gt; 31

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; MSAT-F1 Primer

&lt;400&gt; 102

aataccgcgg aagcttgacc tggaatatcg c

31

&lt;210&gt; 103

&lt;211&gt; 27

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; MSAT-R1 Primer

&lt;400&gt; 103

ataaccgcgg agtccttcag tgtgcat

27

&lt;210&gt; 104

&lt;211&gt; 277

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Nopaline Synthase Promoter Sequence

&lt;300&gt;

&lt;308&gt; Genbank #U09365

&lt;309&gt; 1997-10-17

-62-

<400> 104  
 gagctcgaat ttccccgata gttcaaacat ttggcaataa agtttcttaa gattgaatcc 60  
 tggtgccggt cttgcgatga ttatcatata atttctgttg aattacgtta agcatgtaat 120  
 aattaacatg taatgcatga cgttatttat gagatgggtt tttatgatta ggtcccgca 180  
 attatacatt taatagcgga tagaaaacaa aatatagcgc gcaaactagg ataaattatc 240  
 gcgcgcggtg tcattctatgt tactagatcg ggaattc 277

<210> 105  
 <211> 1812  
 <212> DNA  
 <213> Escherichia coli

<220>  
 <221> CDS  
 <222> (1)...(1812)  
 <223> Beta-Glucuronidase

<300>  
 <308> Genbank #S69414  
 <309> 1994-09-23

<400> 105  
 atg tta cgt cct gta gaa acc cca acc cgt gaa atc aaa aaa ctc gac 48  
 Met Leu Arg Pro Val Glu Thr Pro Thr Arg Glu Ile Lys Lys Leu Asp  
 1 5 10 15

ggc ctg tgg gca ttc agt ctg gat cgc gaa aac tgt gga att gat cag 96  
 Gly Leu Trp Ala Phe Ser Leu Asp Arg Glu Asn Cys Gly Ile Asp Gln  
 20 25 30

cgt tgg tgg gaa agc gcg tta caa gaa agc cgg gca att gct gtg cca 144  
 Arg Trp Trp Glu Ser Ala Leu Gln Glu Ser Arg Ala Ile Ala Val Pro  
 35 40 45

ggc agt ttt aac gat cag ttc gcc gat gca gat att cgt aat tat gcg 192  
 Gly Ser Phe Asn Asp Gln Phe Ala Asp Ala Asp Ile Arg Asn Tyr Ala  
 50 55 60

ggc aac gtc tgg tat cag cgc gaa gtc ttt ata ccg aaa ggt tgg gca 240  
 Gly Asn Val Trp Tyr Gln Arg Glu Val Phe Ile Pro Lys Gly Trp Ala  
 65 70 75 80

ggc cag cgt atc gtg ctg cgt ttc gat gcg gtc act cat tac ggc aaa 288  
 Gly Gln Arg Ile Val Leu Arg Phe Asp Ala Val Thr His Tyr Gly Lys  
 85 90 95

gtg tgg gtc aat aat cag gaa gtg atg gag cat cag ggc ggc tat acg 336  
 Val Trp Val Asn Asn Gln Glu Val Met Glu His Gln Gly Gly Tyr Thr  
 100 105 110

cca ttt gaa gcc gat gtc acg ccg tat gtt att gcc ggc aaa agt gta 384  
 Pro Phe Glu Ala Asp Val Thr Pro Tyr Val Ile Ala Gly Lys Ser Val  
 115 120 125

cgt atc acc gtt tgt gtg aac aac gaa ctg aac tgg cag act atc ccg 432  
 Arg Ile Thr Val Cys Val Asn Asn Glu Leu Asn Trp Gln Thr Ile Pro  
 130 135 140

ccg gga atg gtg att acc gac gaa aac ggc aag aaa aag cag tct tac 480  
 Pro Gly Met Val Ile Thr Asp Glu Asn Gly Lys Lys Lys Gln Ser Tyr  
 145 150 155 160

ttc cat gat ttc ttt aac tat gcc gga atc cat cgc agc gta atg ctc 528  
 Phe His Asp Phe Phe Asn Tyr Ala Gly Ile His Arg Ser Val Met Leu  
 165 170 175

tac acc acg ccg aac acc tgg gtg gac gat atc acc gtg gtg acg cat 576

-63-

Tyr	Thr	Thr	Pro	Asn	Thr	Trp	Val	Asp	Asp	Ile	Thr	Val	Val	Thr	His	
			180					185					190			
gtc	gcg	caa	gac	tgt	aac	cac	gcg	tct	gtt	gac	tgg	cag	gtg	gtg	gcc	624
Val	Ala	Gln	Asp	Cys	Asn	His	Ala	Ser	Val	Asp	Trp	Gln	Val	Val	Ala	
		195					200					205				
aat	ggt	gat	gtc	agc	gtt	gaa	ctg	cgt	gat	gcg	gat	caa	cag	gtg	gtt	672
Asn	Gly	Asp	Val	Ser	Val	Glu	Leu	Arg	Asp	Ala	Asp	Gln	Gln	Val	Val	
	210					215					220					
gca	act	gga	caa	ggc	act	agc	ggg	act	ttg	caa	gtg	gtg	aat	ccg	cac	720
Ala	Thr	Gly	Gln	Gly	Thr	Ser	Gly	Thr	Leu	Gln	Val	Val	Asn	Pro	His	
	225				230					235					240	
ctc	tgg	caa	ccg	ggt	gaa	ggt	tat	ctc	tat	gaa	ctg	tgc	gtc	aca	gcc	768
Leu	Trp	Gln	Pro	Gly	Glu	Gly	Tyr	Leu	Tyr	Glu	Leu	Cys	Val	Thr	Ala	
				245					250					255		
aaa	agc	cag	aca	gag	tgt	gat	atc	tac	ccg	ctt	cgc	gtc	ggc	atc	cgg	816
Lys	Ser	Gln	Thr	Glu	Cys	Asp	Ile	Tyr	Pro	Leu	Arg	Val	Gly	Ile	Arg	
			260					265					270			
tca	gtg	gca	gtg	aag	ggc	gaa	cag	ttc	ctg	att	aac	cac	aaa	ccg	ttc	864
Ser	Val	Ala	Val	Lys	Gly	Glu	Gln	Phe	Leu	Ile	Asn	His	Lys	Pro	Phe	
		275					280					285				
tac	ttt	act	ggc	ttt	ggt	cgt	cat	gaa	gat	gcg	gac	ttg	cgt	ggc	aaa	912
Tyr	Phe	Thr	Gly	Phe	Gly	Arg	His	Glu	Asp	Ala	Asp	Leu	Arg	Gly	Lys	
	290					295					300					
gga	ttc	gat	aac	gtg	ctg	atg	gtg	cac	gac	cac	gca	tta	atg	gac	tgg	960
Gly	Phe	Asp	Asn	Val	Leu	Met	Val	His	Asp	His	Ala	Leu	Met	Asp	Trp	
	305				310					315					320	
att	ggg	gcc	aac	tcc	tac	cgt	acc	tcg	cat	tac	cct	tac	gct	gaa	gag	1008
Ile	Gly	Ala	Asn	Ser	Tyr	Arg	Thr	Ser	His	Tyr	Pro	Tyr	Ala	Glu	Glu	
				325					330					335		
atg	ctc	gac	tgg	gca	gat	gaa	cat	ggc	atc	gtg	gtg	att	gat	gaa	act	1056
Met	Leu	Asp	Trp	Ala	Asp	Glu	His	Gly	Ile	Val	Val	Ile	Asp	Glu	Thr	
			340					345					350			
gct	gct	gtc	ggc	ttt	aac	ctc	tct	tta	ggc	att	ggt	ttc	gaa	gcg	ggc	1104
Ala	Ala	Val	Gly	Phe	Asn	Leu	Ser	Leu	Gly	Ile	Gly	Phe	Glu	Ala	Gly	
		355					360					365				
aac	aag	ccg	aaa	gaa	ctg	tac	agc	gaa	gag	gca	gtc	aac	ggg	gaa	act	1152
Asn	Lys	Pro	Lys	Glu	Leu	Tyr	Ser	Glu	Glu	Ala	Val	Asn	Gly	Glu	Thr	
	370					375					380					
cag	caa	gcg	cac	tta	cag	gcg	att	aaa	gag	ctg	ata	gcg	cgt	gac	aaa	1200
Gln	Gln	Ala	His	Leu	Gln	Ala	Ile	Lys	Glu	Leu	Ile	Ala	Arg	Asp	Lys	
				390						395					400	
aac	cac	cca	agc	gtg	gtg	atg	tgg	agt	att	gcc	aac	gaa	ccg	gat	acc	1248
Asn	His	Pro	Ser	Val	Val	Met	Trp	Ser	Ile	Ala	Asn	Glu	Pro	Asp	Thr	
				405					410					415		
cgt	ccg	caa	ggt	gca	cgg	gaa	tat	ttc	gcg	cca	ctg	gcg	gaa	gca	acg	1296
Arg	Pro	Gln	Gly	Ala	Arg	Glu	Tyr	Phe	Ala	Pro	Leu	Ala	Glu	Ala	Thr	
			420					425					430			
cgt	aaa	ctc	gac	ccg	acg	cgt	ccg	atc	acc	tgc	gtc	aat	gta	atg	ttc	1344
Arg	Lys	Leu	Asp	Pro	Thr	Arg	Pro	Ile	Thr	Cys	Val	Asn	Val	Met	Phe	
		435					440					445				

-64-

tgc	gac	gct	cac	acc	gat	acc	atc	agc	gat	ctc	ttt	gat	gtg	ctg	tgc	1392
Cys	Asp	Ala	His	Thr	Asp	Thr	Ile	Ser	Asp	Leu	Phe	Asp	Val	Leu	Cys	
	450					455					460					
ctg	aac	cgt	tat	tac	gga	tgg	tat	gtc	caa	agc	ggc	gat	ttg	gaa	acg	1440
Leu	Asn	Arg	Tyr	Tyr	Gly	Trp	Tyr	Val	Gln	Ser	Gly	Asp	Leu	Glu	Thr	
465					470					475					480	
gca	gag	aag	gta	ctg	gaa	aaa	gaa	ctt	ctg	gcc	tgg	cag	gag	aaa	ctg	1488
Ala	Glu	Lys	Val	Leu	Glu	Lys	Glu	Leu	Leu	Ala	Trp	Gln	Glu	Lys	Leu	
				485					490					495		
cat	cag	ccg	att	atc	atc	acc	gaa	tac	ggc	gtg	gat	acg	tta	gcc	ggg	1536
His	Gln	Pro	Ile	Ile	Ile	Thr	Glu	Tyr	Gly	Val	Asp	Thr		Ala	Gly	
			500					505					510			
ctg	cac	tca	atg	tac	acc	gac	atg	tgg	agt	gaa	gag	tat	cag	tgt	gca	1584
Leu	His	Ser	Met	Tyr	Thr	Asp	Met	Trp	Ser	Glu	Glu	Tyr	Gln	Cys	Ala	
		515					520					525				
tgg	ctg	gat	atg	tat	cac	cgc	gtc	ttt	gat	cgc	gtc	agc	gcc	gtc	gtc	1632
Trp	Leu	Asp	Met	Tyr	His	Arg	Val	Phe	Asp	Arg	Val	Ser	Ala	Val	Val	
	530					535					540					
ggg	gaa	cag	gta	tgg	aat	ttc	gcc	gat	ttt	gcg	acc	tcg	caa	ggc	ata	1680
Gly	Glu	Gln	Val	Trp	Asn	Phe	Ala	Asp	Phe	Ala	Thr	Ser	Gln	Gly	Ile	
545					550					555					560	
ttg	cgc	gtt	ggc	ggg	aac	aag	aaa	ggg	atc	ttc	act	cgc	gac	cgc	aaa	1728
Leu	Arg	Val	Gly	Gly	Asn	Lys	Lys	Gly	Ile	Phe	Thr	Arg	Asp	Arg	Lys	
			565					570						575		
ccg	aag	tcg	gcg	gct	ttt	ctg	ctg	caa	aaa	cgc	tgg	act	ggc	atg	aac	1776
Pro	Lys	Ser	Ala	Ala	Phe	Leu	Leu	Gln	Lys	Arg	Trp	Thr	Gly	Met	Asn	
			580					585					590			
ttc	ggg	gaa	aaa	ccg	cag	cag	gga	ggc	aaa	caa	tga					1812
Phe	Gly	Glu	Lys	Pro	Gln	Gln	Gly	Gly	Lys	Gln	*					
		595					600									

&lt;210&gt; 106

&lt;211&gt; 603

&lt;212&gt; PRT

&lt;213&gt; Escherichia coli

&lt;300&gt;

&lt;308&gt; Genbank #S69414

&lt;309&gt; 1994-09-23

&lt;400&gt; 106

Met	Leu	Arg	Pro	Val	Glu	Thr	Pro	Thr	Arg	Glu	Ile	Lys	Lys	Leu	Asp	
1				5					10					15		
Gly	Leu	Trp	Ala	Phe	Ser	Leu	Asp	Arg	Glu	Asn	Cys	Gly	Ile	Asp	Gln	
			20					25					30			
Arg	Trp	Trp	Glu	Ser	Ala	Leu	Gln	Glu	Ser	Arg	Ala	Ile	Ala	Val	Pro	
			35				40					45				
Gly	Ser	Phe	Asn	Asp	Gln	Phe	Ala	Asp	Ala	Asp	Ile	Arg	Asn	Tyr	Ala	
	50					55					60					
Gly	Asn	Val	Trp	Tyr	Gln	Arg	Glu	Val	Phe	Ile	Pro	Lys	Gly	Trp	Ala	
65					70					75					80	
Gly	Gln	Arg	Ile	Val	Leu	Arg	Phe	Asp	Ala	Val	Thr	His	Tyr	Gly	Lys	
			85					90					95			
Val	Trp	Val	Asn	Asn	Gln	Glu	Val	Met	Glu	His	Gln	Gly	Gly	Tyr	Thr	
			100					105					110			
Pro	Phe	Glu	Ala	Asp	Val	Thr	Pro	Tyr	Val	Ile	Ala	Gly	Lys	Ser	Val	
		115					120					125				

-65-

Arg	Ile	Thr	Val	Cys	Val	Asn	Asn	Glu	Leu	Asn	Trp	Gln	Thr	Ile	Pro
130						135					140				
Pro	Gly	Met	Val	Ile	Thr	Asp	Glu	Asn	Gly	Lys	Lys	Lys	Gln	Ser	Tyr
145				150						155					160
Phe	His	Asp	Phe	Phe	Asn	Tyr	Ala	Gly	Ile	His	Arg	Ser	Val	Met	Leu
				165					170					175	
Tyr	Thr	Thr	Pro	Asn	Thr	Trp	Val	Asp	Asp	Ile	Thr	Val	Val	Thr	His
			180					185					190		
Val	Ala	Gln	Asp	Cys	Asn	His	Ala	Ser	Val	Asp	Trp	Gln	Val	Val	Ala
		195					200					205			
Asn	Gly	Asp	Val	Ser	Val	Glu	Leu	Arg	Asp	Ala	Asp	Gln	Gln	Val	Val
	210					215					220				
Ala	Thr	Gly	Gln	Gly	Thr	Ser	Gly	Thr	Leu	Gln	Val	Val	Asn	Pro	His
225				230						235					240
Leu	Trp	Gln	Pro	Gly	Glu	Gly	Tyr	Leu	Tyr	Glu	Leu	Cys	Val	Thr	Ala
				245					250					255	
Lys	Ser	Gln	Thr	Glu	Cys	Asp	Ile	Tyr	Pro	Leu	Arg	Val	Gly	Ile	Arg
			260					265					270		
Ser	Val	Ala	Val	Lys	Gly	Glu	Gln	Phe	Leu	Ile	Asn	His	Lys	Pro	Phe
		275				280					285				
Tyr	Phe	Thr	Gly	Phe	Gly	Arg	His	Glu	Asp	Ala	Asp	Leu	Arg	Gly	Lys
	290					295				300					
Gly	Phe	Asp	Asn	Val	Leu	Met	Val	His	Asp	His	Ala	Leu	Met	Asp	Trp
305					310					315					320
Ile	Gly	Ala	Asn	Ser	Tyr	Arg	Thr	Ser	His	Tyr	Pro	Tyr	Ala	Glu	Glu
			325						330					335	
Met	Leu	Asp	Trp	Ala	Asp	Glu	His	Gly	Ile	Val	Val	Ile	Asp	Glu	Thr
			340					345					350		
Ala	Ala	Val	Gly	Phe	Asn	Leu	Ser	Leu	Gly	Ile	Gly	Phe	Glu	Ala	Gly
		355					360					365			
Asn	Lys	Pro	Lys	Glu	Leu	Tyr	Ser	Glu	Glu	Ala	Val	Asn	Gly	Glu	Thr
	370					375					380				
Gln	Gln	Ala	His	Leu	Gln	Ala	Ile	Lys	Glu	Leu	Ile	Ala	Arg	Asp	Lys
385				390						395					400
Asn	His	Pro	Ser	Val	Val	Met	Trp	Ser	Ile	Ala	Asn	Glu	Pro	Asp	Thr
			405						410					415	
Arg	Pro	Gln	Gly	Ala	Arg	Glu	Tyr	Phe	Ala	Pro	Leu	Ala	Glu	Ala	Thr
			420					425					430		
Arg	Lys	Leu	Asp	Pro	Thr	Arg	Pro	Ile	Thr	Cys	Val	Asn	Val	Met	Phe
		435					440					445			
Cys	Asp	Ala	His	Thr	Asp	Thr	Ile	Ser	Asp	Leu	Phe	Asp	Val	Leu	Cys
	450					455					460				
Leu	Asn	Arg	Tyr	Tyr	Gly	Trp	Tyr	Val	Gln	Ser	Gly	Asp	Leu	Glu	Thr
465					470					475					480
Ala	Glu	Lys	Val	Leu	Glu	Lys	Glu	Leu	Leu	Ala	Trp	Gln	Glu	Lys	Leu
				485					490					495	
His	Gln	Pro	Ile	Ile	Ile	Thr	Glu	Tyr	Gly	Val	Asp	Thr	Leu	Ala	Gly
		500						505					510		
Leu	His	Ser	Met	Tyr	Thr	Asp	Met	Trp	Ser	Glu	Glu	Tyr	Gln	Cys	Ala
	515						520					525			
Trp	Leu	Asp	Met	Tyr	His	Arg	Val	Phe	Asp	Arg	Val	Ser	Ala	Val	Val
	530					535					540				
Gly	Glu	Gln	Val	Trp	Asn	Phe	Ala	Asp	Phe	Ala	Thr	Ser	Gln	Gly	Ile
545					550					555					560
Leu	Arg	Val	Gly	Gly	Asn	Lys	Lys	Gly	Ile	Phe	Thr	Arg	Asp	Arg	Lys
				565					570					575	
Pro	Lys	Ser	Ala	Ala	Phe	Leu	Leu	Gln	Lys	Arg	Trp	Thr	Gly	Met	Asn
			580					585					590		
Phe	Gly	Glu	Lys	Pro	Gln	Gln	Gly	Gly	Lys	Gln					
		595					600								

&lt;210&gt; 107

&lt;211&gt; 277

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

-66-

&lt;220&gt;

&lt;223&gt; Nopaline Synthase Terminator Sequence

&lt;300&gt;

&lt;308&gt; U09365

&lt;309&gt; 1995-10-17

&lt;400&gt; 107

gagctcgaat	ttccccgatc	gttcaaacat	ttggcaataa	agtttcttaa	gattgaatcc	60
tggtgccggt	cttgcgatga	ttatcatata	atttctgttg	aattacgtta	agcatgtaat	120
aattaacatg	taatgcatga	cgttatttat	gagatgggtt	tttatgatta	gagtccccga	180
attatacatt	taatacgcga	tagaaaacaa	aatatagcgc	gcaaaactagg	ataaattatc	240
gcgcgcggtg	tcattctatgt	tactagatcg	ggaattc			277

&lt;210&gt; 108

&lt;211&gt; 3451

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> HindIII Fragment containing the beta-glucuronidase coding sequence, the rDNA intergenic spacer, and the Mast1 sequence

&lt;400&gt; 108

aagcttgacc	tggaatatcg	cgagtaaact	gaaaatcacg	gaaaatgaga	aatacacact	60
ttaggacgtg	aaatatggcg	aggaaaactg	aaaaagggtg	aaaatttaga	aatgtccact	120
gtaggacgtg	gaatatggca	agaaaactga	aaatcatgga	aaatgagaaa	catccacttg	180
acgacttgaa	aaatgacgaa	atcactaaaa	aacgtgaaaa	atgagaaatg	cacactgaag	240
gactccgcgg	gaattcgatt	gtgctagcca	atgtttaaca	agatgtcaag	cacaatgaat	300
gttggtggtt	gggtggtcgtg	gctggcgggtg	gtggaaaatt	gcgggtggttc	gagcggtagt	360
gatcggcgat	gggtggtggtt	tcgagcgggtg	tttgatatcg	gaatcactta	tggtggttgt	420
cacaatggag	gtgcgtcatg	gttattgggtg	gttggtcatc	tatatatttt	tataataata	480
ttaaagtattt	tacctatttt	ttacatatatt	tttattaaat	ttatgcattg	tttgtatttt	540
taaaatagttt	ttatcgtact	tgttttataa	aatatttttat	tatttttatgt	gttatattat	600
tacttgatgt	attggaaatt	ttctccattg	ttttttctat	atttataata	attttcttat	660
ttttttttgt	tttatttatgt	attttttctg	tttataataa	atatttatta	aaaaaaatat	720
tattttttgta	aaatatatca	tttacaatgt	ttaaaagtca	tttgtgaata	tattagctaa	780
gttgtaacttc	tttttgtgca	tttggtgttg	tacatgtcta	ttatgattct	ctggccaaa	840
catgtctact	cctgtcactt	gggttttttt	ttttaagaca	taatcactag	tgattatata	900
tagactgaag	gcgggaaacg	acaatctgat	catgagcgga	gaattaaagg	agtcacgtta	960
tgacccccgc	cgatgacgcg	ggacaagccg	ttttacgttt	ggaactgaca	gaaccgcaac	1020
gttgaggag	ccactcagcc	gcgggtttct	ggagtttaaat	gagctaagca	catacgtcag	1080
aaaccattat	tgcgcggttca	aaagtcgcct	aaggctacta	tcagctagca	aatatttctt	1140
gtcaaaaatg	ctccactgac	gttccataaa	ttcccctcgg	tatccaatta	gagtcctata	1200
ttcactctca	atccaaataa	tctgcaccgg	atctcgagat	cgaattcccg	cggcgcgcaa	1260
ttcactagt	gatccccggg	tacggtcagt	cccttatgtt	acgtcctgta	gaaaccccaa	1320
cccgatgaat	caaaaaactc	gacggcctgt	gggcattcag	tctggatcgc	gaaaactgtg	1380
gaattgagca	gcgttggtgg	gaaagcgcgt	tacaagaaag	ccgggcaatt	gctgtgccag	1440
gcagttttta	cgatcagttc	gccgatgcag	atattcgtaa	ttatgtgggc	aacgtctggt	1500
atcagcgca	agtctttata	ccgaaagggt	gggcaggcca	gcgtatcgtg	ctgcgtttcg	1560
atgcggtcac	tcattacggc	aaagtgtggg	tcaataatca	ggaagtgatg	gagcatcagg	1620
gcggctatac	gccatttgaa	gccgatgtca	cgccgtatgt	tattgcgggg	aaaagtgtac	1680
gtatcacagt	ttgtgtgaac	aacgaactga	actggcagac	tatcccgccg	ggaatggtga	1740
ttaccgacga	aaacggcaag	aaaaagcagt	cttacttcca	tgatttcttt	aactacgccg	1800
ggatccatcg	cagcgtaatg	ctctacacca	cgccgaacac	ctgggtggac	gatatacccg	1860
tggtgacgca	tgctgcgcaa	gactgtaaac	acgcgtctgt	tgactggcag	gtggtggcca	1920
atggtgatgt	cagcgttgaa	ctgcgtgatg	cggatcaaca	ggtggttgca	actggacaag	1980
gcaccagcgg	gactttgcaa	gtggtgaatc	cgcacctctg	gcaaccgggt	gaagggttatc	2040
tctatgaact	gtacgtcaca	gccaaaagcc	agacagagt	tgatatctac	ccgctgcgcg	2100
tcggcatccg	gtcagtgcca	gtgaaggcgc	aaacagttct	gatcaaccac	aaacggttct	2160
actttactgg	ctttggccgt	catgaagatg	cggatttgcg	cgccaaagga	ttcgataacg	2220
tgctgatggt	gcacgatcac	gcattaatgg	actggattgg	ggccaactcc	taccgtacct	2280
cgcattaccc	ttacgctgaa	gagatgctcg	actgggcaga	tgaacatggc	atcggtggtga	2340
ttgatgaaac	tcagctgtc	ggctttaacc	tctctttagg	cattgggttc	gaagcgggca	2400
acaagccgaa	agaactgtac	agcgaagagg	cagtcaacgg	ggaaactcag	caggcgcact	2460
tacaggcgat	taaagagctg	atagcgcgtg	acaaaaacca	cccaagcgtg	gtgatgtgga	2520



-67-

gtattgcca	cgaaccggat	acccgtccgc	aaggtgcacg	ggaatatttc	gcgccactgg	2580
cggaagcaac	gcgtaaac	gatccgacgc	gtccgatcac	ctgcgtcaat	gtaatgttct	2640
gcgacgtca	caccgatacc	atcagcgatc	tctttgatgt	gctgtgcctg	aaccgttatt	2700
acggttggt	tgtccaaagc	ggcgatttgg	aaacggcaga	gaaggtactg	gaaaaagaac	2760
ttctggcctg	gcaggagaaa	ctgcatcagc	cgattatcat	caccgaatac	ggcgtggata	2820
cgttagccgg	gctgcaactc	atgtacaccg	acatgtggag	tgaagagtat	cagtgtgcat	2880
ggctggatat	gtatcaccgc	gtctttgatc	gcgtcagcgc	cgctcgtcgg	gaacaggtat	2940
ggaatttcgc	cgatttttgcg	acctcgcaag	gcattattgcg	cgttggcggg	aacaagaagg	3000
ggatcttcac	ccgcgaccgc	aaaccgaagt	cgggcgcttt	tctgctgcaa	aaacgtgga	3060
ctggcatgaa	cttcggtgaa	aaaccgcagc	agggaggcaa	acaatgaatc	aacaactctc	3120
ctggcgcacc	atcgctcggt	acagcctcgg	gaattgcgta	ccgagctcga	atttccccga	3180
tcgttcaaac	atctggcaat	aaagtctctt	aagattgaat	cctgttgccg	gtcttgcat	3240
gattatcata	taatttctgt	tgaattacgt	taagcatgta	ataattaaca	tgtaatgcat	3300
gacgttattt	atgagatggg	tttttatgat	tagagtcgcc	caattataca	tttaatacgc	3360
gatagaaaac	aaaatatagc	gcgcaaac	tgataaatta	tcgcgcgcgg	tgcatctat	3420
gttactagat	cggaattcgc	atatcaagct	t			3451

&lt;210&gt; 109

&lt;211&gt; 14627

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pAg11a Plasmid

&lt;400&gt; 109

catgccaac	acaggggttc	cctcggggtc	aaagtacttt	gatccaaccc	ctccgctgct	60
atagtgcagt	cggtcttctga	cggtcagtcg	agccgtcttc	tgaaaacgac	atgtcgcaca	120
agtcctaagt	tacgcgacag	gctgcgcgcc	tgcccttttc	ctggcggttt	cttgctcgct	180
gttttagtcg	cataaagtag	aataacttgcg	actagaaccg	gagacattac	gccatgaaca	240
agagcgcgcg	cgctggcctg	ctgggctatg	cccgcgtcag	caccgacgac	caggacttga	300
ccaaccaacg	ggcggaaactg	cacgcggcgc	gctgcacca	gctgttttcc	gagaagatca	360
ccggcaccag	gcgcgaccgc	ccggagctgg	caggatgct	tgaccaccta	cgccctggcg	420
acgttgtgac	agtgaccagg	ctagaccgcc	tggcccgag	caccgcgcac	ctactggaca	480
ttgccagcgc	catccaggag	gcccgcgcgc	gcctgcgtag	cctggcagag	ccgtgggccc	540
acaccaccac	cccgccggcg	cgcattggtgt	tgaccgtgtt	cgccggcatt	gccgagttcg	600
agcgttccct	aatcatcgac	cgcaccggga	gcgggcgcg	ggccgccaag	gcccagggcg	660
tgaagtttgg	cccccgccct	accctcaccc	cggcacagat	cgcgacgcgc	cgcgagctga	720
tcgacaggga	aggcgcgacc	gtgaaagagg	cggtcgact	gcttggcgtg	catcgctgga	780
ccctgtaccg	cgcatcttag	cgacgcgagg	aagtgcagcc	caccgaggcc	aggcggcgcg	840
gtgccttccg	tgaggacgca	ttgaccgagg	ccgacgcctt	ggcgccgcgc	gagaatgaac	900
gccaaagagga	acaagcatga	aaccgcacca	ggagggccag	gacgaaccgt	ttttcattac	960
cgaagagatc	gaggcgagga	tgatcgcggc	cggttactgt	ttcgagccgc	ccgcgcacgt	1020
ctcaaccgtg	cggttgcgtg	aaatcctggc	cggtttgtct	gatgccaagc	tggcgccctg	1080
gcccggccagc	ttggccgctg	aagaaaaccga	gcgcgcgcgt	ctaaaaaggt	gatgtgtatt	1140
tgagtaaaac	agcttgcgtc	atgcgggtcgc	tgcttatatg	atgcgatgag	taataaaaca	1200
aatacgcaag	gggaacgcct	gaaggttatc	gctgtactta	accagaaagg	cggttcaggc	1260
aagacgacca	tcgcaaccca	tctagcccg	gccttgcaac	tcgcccgggc	cgatgttctg	1320
ttagtcgatt	ccgatcccca	gggcagtgcc	cgcgattggg	cgcccggtgcg	ggaagatcaa	1380
ccgctaaccg	ttgtcggcat	cgaccgcgcg	acgattgacc	gcgacgtgaa	ggccatccgg	1440
cggcgcgact	tcgtagtgt	cgacggagcg	cccaggcgcg	cggaacttgg	tgtgtccgcg	1500
atcaaggcag	ccgacttcgt	gctgattccg	gtgcagccaa	gcccttacga	catatggggc	1560
accgcgcgac	tggtggagct	ggttaagcag	cgcattgagg	tcacggatgg	aaggctacaa	1620
gggcgccttg	tcgtgtcgcg	ggcgatcaaa	ggcacgcgca	tcggcggtga	ggttgccgag	1680
gcgctggccg	ggtaacgagct	gcccattctt	gagtcgctga	tcacgcagcg	cgtgagctac	1740
ccaggcactg	ccgcgcgcgc	cacaaccggt	cttgaatcag	aaccgcaggg	cgacgctgcc	1800
cgcgaggtcc	aggcgctggc	cgctgaaat	aaatcaaaac	tcatttgagt	taatgaggt	1860
aagagaaaaa	gagcaaaaagc	acaaaacacg	taagtgcgcg	ccgtccgagc	gcacgcgagc	1920
gcaaggctgc	aacgttggcc	agcctggcag	acacgcgcgc	catgaagcgg	gtcaactttc	1980
agttgcccgc	ggaggatcac	accaagctga	agatgtacgc	ggtacgcca	ggcaagacca	2040
ttaccgagct	gctatctgaa	tacatcgcg	agctaccaga	gtaaatgagc	aaatgaataa	2100
atgagtagat	gaatttttagc	ggctaaagga	ggcgcatggg	aaaatcaaga	acaaccaggc	2160
accgacgcgc	tgggaatgcc	catgtgtgga	ggaacgggcg	gttggccagg	cgtaagcggc	2220
tgggttgctc	gccggccctg	caatggcact	ggaaccccca	agcccaggga	atcggcgtga	2280
cggtcgcaaa	ccatccggcc	cggtacaaat	cggcgcgcgc	ctgggtgatg	acctggtgga	2340
gaagttgaag	gcccgcgcag	ccgcccagcg	gcaacgcctc	gaggcagaag	cacgccccgg	2400
tgaatcgtgg	caagcggcgc	ctgatcgaat	ccgcaaaaga	tcccggcaac	cgccggcagc	2460

cgggtgcgcgcg	tcgatttagga	agccgcgccaa	gggcgcacgag	caaccagatt	ttttcggttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tcgcagcatc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	ggtgatccgc	tacgagcttc	cagacgggca	2640
cgtagaggtt	tccgcagggc	cggccggcat	ggccagtggt	tgggattacg	acctgggtact	2700
gatggcgggtt	tcccatctaa	ccgaatccat	gaaccgatac	cgggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaaag	cagaaagacg	acctggtaga	aacctgcatt	cgggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggccgcctg	gtgacggtat	ccgaggggtg	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaacc	gggcggcccg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggacgtgct	3060
gacgggttcac	cccgattact	ttttgatcga	tcccggcatc	ggccgttttc	tctaccgctt	3120
ggcagccgcg	gcccagggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gcccggagagt	tcaagaagtt	ctgtttcacc	gtgcgcgaagc	tgatcgggtc	3240
aaatgacctg	ccggagtagc	atttgaaggc	ggaggcgggg	caggctggcc	cgatcctagt	3300
catgcgctac	cgcaacctga	tcgagggcga	agcatccgcc	ggttcctaata	gtacggagca	3360
gatgctaggg	caaattgccc	tagcagggga	aaaaggctcg	aaaggctctct	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaaaccgt	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	actgtaagt	actgatataa	aagagaaaaa	3540
aggcgatttt	tccgcctaaa	actctttaaa	acttattaaa	actcttaaaa	cccgcctggc	3600
ctgtgcataa	ctgtctggcc	agcgcacagc	cgaagagctg	caaaaagcgc	ctacccctcg	3660
gtcgctgcgc	tccctacgcc	cgcctcgctt	cgcctcgccg	atcgcgcccg	ctggccgctc	3720
aaaaatggct	ggcctaaggc	caggcaatct	accaggcgcc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	cggcgcccac	atcaaggcac	cctgcctcgc	gcgtttcggt	gatgacgggtg	3840
aaaacctctg	acacatgcag	ctcccggaga	cggtcacagc	ttgtctgtaa	gcggatggcg	3900
ggagcagaca	agcccgtcag	ggcgcgtcag	cgggtgttgg	cgggtgtcgg	ggcgcagcca	3960
tgaccagctg	acgtagcgat	agcggagtg	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcggtg	tgaaataccg	cacagatgcg	taaggagaaa	4080
ataccgcac	aggcgctctt	cgccttcttc	gtcactgac	tcgctgcgct	cggctcggtc	4140
gctgcggcga	gcgggtatcag	ctcactcaaa	ggcggttaata	cggttatcca	cagaatcagg	4200
ggataacgca	ggaaaagaaca	tgtagagcaa	aaggccagga	aaggccagga	acctgatccg	4260
ggcgcggttg	ctggcggtttt	tccataggct	ccgccccctt	gacgagcatc	acaaaaatcg	4320
acgctcaagt	caagaggtggc	gaaacccgac	aggactataa	agataccagg	cgtttcccc	4380
tggaaagctc	tcctgtgcct	ctcctgttcc	gacctggccg	cttaccggga	acctgtccgc	4440
ctttctccct	tcgggaagcg	tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	4500
ggtgtaggct	gttcgctcca	agctgggctg	tgtagcagaa	ccccccgttc	agccccagcc	4560
ctgcgcctta	tcoggttaact	atcgtcttga	gtccaacccg	gtaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tggtggccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4740
tctgctgaag	ccagttacct	tcggaaaaag	agttggtagc	tcttgatccg	gcaacaaaa	4800
caccgctggg	agcgggtgggt	tttttggttg	caagcagcag	attaacgcga	gaaaaaaagg	4860
atctcaagaa	gatcctttga	tctttctctc	ggggtctgac	gctcagtgga	acgaaaaact	4920
acgttaaggg	atttttggtc	tgcatcttag	gtactaaaac	aattcatoca	gtaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gtcgcacata	5040
ctgttcttcc	ccgataatct	ccctgatcga	cgggacgcag	aaggcaatgt	cataccactt	5100
gtccgccttg	ccgcttctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgctg	tctcccagg	cgccgtggga	aaagacaagt	tccctcttcg	gcttttccgt	5220
ctttaaaaaa	tcatacagct	cgcgcgggat	tttaaatgga	gtgtcttctt	ccagtttttc	5280
gcaatccaca	tcggccagat	cgttattcag	taagtaattc	aattcggcta	agcggctgtc	5340
taagctattc	gtatagggac	aatccgatat	gtcgatggag	tgaaagagcc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggtc	ttgttcacat	tcatactctt	ccgagcaaac	5460
gacgccatcg	gcctcactca	tgagcagatt	gtccagcca	tcatgcccgt	caaagtcag	5520
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5580
atcataagtg	gtccctttat	accggctgtc	cgtcattttt	aaatataggt	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgcagcggta	5700
tttttcgata	agttttttca	attccgggtg	tattctcatt	ttagccattt	attatttctt	5760
tcctcttttc	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tccttgcttt	ctaaaacctt	aaataccaga	aaacagcttt	ttcaaagttg	5880
ttttcaaagt	tggcgtataa	catagtatcg	acggagccga	ttttgaaacc	gcgggtgatca	5940
caggcagcaa	cgctctgtca	tcgttacaat	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaac	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaac	6060
tctgccgcct	tacaacggct	tcctccgtga	cgcgcgtccg	gactgatggg	ctgcctgtat	6120
cgagtgggtg	ttttgtgccc	agctgcccgt	cggggagctg	ttggctgggt	ggtggcagga	6180
tatatgtggt	tgtaacaaa	ttgacgctta	gacaacctaa	taacacattg	cggacgtttt	6240
taatgtactg	aattaacgcc	gaattaattc	gggggatctg	gatttttagta	ctggattttg	6300
gttttaggaa	ttagaaattt	tattgataga	agttttttac	aaatacaaat	acataactag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctatagg	aaccctaatt	cccttatctg	6420
ggaactactc	acacattatt	atggagaaac	tcgagtcaaa	tctcggtgac	gggcaggacc	6480

ggacggggcg	gtaccggcag	gctgaagtcc	agctgccaga	aacccacgtc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgcctcgtgc	6600
atgcgcacgc	tcgggtcggt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660
gcctccaggg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtccegt	ccgctgggtg	6720
cgggggggaga	cgtacacggg	cgactcggcc	gtccagtcgt	aggcgtttgc	tgccctccag	6780
gggcccgcgt	aggcgatgcc	ggcgacctcg	ccgtccacct	cggcgacgag	ccagggatag	6840
cgctcccgcga	gacggacgag	gtcgtccgtc	cactcctgcg	gttcctgcgg	ctcgggtacg	6900
aagttgaccg	tgcttgtctc	gatgtagtgg	ttgacgatgg	tgacagaccg	cggcatgtcc	6960
gcctcgggtg	cacggcggat	gtcggccggg	cgctgctctg	ggctcatggg	agactcgaga	7020
gagatagatt	tgtagagaga	gactgggtgat	ttcagcgtgt	cctctccaaa	tgaatgaac	7080
ttccttata	agagggaagg	cttgcggaagg	atagtgggat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaacgt	cttctttttc	7200
cacgatgtct	ctcgtgggtg	ggggtccatc	tttgggacca	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttcttttat	cgcaatgatg	gcatttgtag	gtgccacctt	ccctttctac	7320
tgctcttttg	atgaagtgc	agatagctgg	cgatgtgaat	ccgaggaggt	ttcccgatag	7380
taccttttgt	tgaagagtct	caatagccct	ttggtcttct	gagactgtat	ctttgatatt	7440
cttggagtag	acgagagtgt	cggtgtccac	catgttatca	catcaatcca	cttgctttga	7500
agacgtgggt	ggaacgtctt	ctttttccac	cgctgctctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcttgaac	gatagccttt	cctttatcgc	aatgatggca	7620
tttgtagggt	ccaccttcc	tttctactgt	ccttttgatg	aagtgcagag	tagctggggg	7680
atggaatccg	aggaggtttc	ccgatattac	ctcttgttga	aaagtctcaa	tagccctttg	7740
gtcttctgag	actgtatctt	tgatattctt	ggagtgcagc	agagtgtcgt	gtccaccat	7800
gttggcaagc	tgctctagcc	aatacgcaaa	ccgcctctcc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcagacag	gtttcccgac	tggaaagcgg	gcagtgcgag	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcacc	caggctttac	actttatgct	tcgggctcgt	7980
atgttgtgtg	gaattgtgag	cggataacaa	tttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agccttgact	agaggttoga	cggtatacag	acatgataag	atacatgtat	8100
gagtttggac	aaaccacaac	tagaatgcag	tgaaaaaaat	gctttatttg	tgaattttgt	8160
gatgtctatt	ctttatttgt	aaccattata	agctgcaata	aacaagttgg	gggtgggcgaa	8220
gaactccagc	atgagatccc	cgcgtcggag	gatcatccag	ccggcgtccc	ggaaaacgat	8280
tccgaagccc	aacctttcat	agaaggcgcc	gggtggaatc	aaatctcgta	gcacgtgtca	8340
gtcctgctcc	tcggccacga	agtgcacgca	gttgccggcc	gggtcgcgca	gggcgaactc	8400
ccgccccac	ggctgctcgc	cgatctcggt	cagctgcgtc	ccggaggcgt	cccacacca	8460
cgtggacacg	acctccgacc	actcggcgta	catgctgtcc	aggccgcgca	cccacacca	8520
ggccagggtg	ttgtccggca	ccacctgggt	ctggaccgcg	ctgatgaaca	gggtcacgtc	8580
gtcccgacc	acaccggcga	agtcgtctct	cacgaagtcc	cgggagaacc	cgaccgggtc	8640
ggccagacac	tcgaccgctc	cggcgacgtc	gcgcgcgggt	agcaccggaa	cggcactggg	8700
caacttggcc	atggatccag	atttcgctca	agttagtata	aaaaagcagg	cttcaatcct	8760
gcaggaattc	gatcgacact	ctcgtctact	ccaagaatat	caaagataca	gtctcagaag	8820
accaaaggcc	tattgagact	tttcaacaaa	gggttaatat	gggaaacctc	ctcggattcc	8880
attgcccagc	tatctgtcac	ttcatcaaaa	ggacagttaga	aaagggaagg	ggcacctaca	8940
aatgccatca	ttcgataaaa	ggaaaggcta	tcgttcaaga	tgccctctgcc	gcacgtgggt	9000
ccaaagatgg	acccccaccc	acgaggagca	tcgtggaaaa	agaagacgtt	ccaaccacgt	9060
cttcaaagca	agtggattga	tgtgataaca	tggtggagca	cgacactctc	gtctactcca	9120
agaatatcaa	agatacagtc	tcagaagacc	aaagggtat	tgagactttt	caacaaaggg	9180
taatatcggg	aaacctctc	ggattccatt	gcccagctat	ctgtcacttc	atcaaaagga	9240
cagtagaaaa	ggaagggtgg	acctacaaat	gccatcattg	cgataaagga	aaggctatcg	9300
ttcaagatgc	ctctgccgac	agtggtccca	aagatggacc	cccaccacag	aggagcatcg	9360
tggaaaaaga	agacgttcca	accacgtctt	caaagcaagt	ggattgatgt	gatatctcca	9420
ctgacgtaag	ggatgacgca	caatcccact	atccttcgca	agaccttcc	ctatataagg	9480
aagttcattt	catttgagga	ggacacgctg	aaatcaccag	tctctctcta	caaactctatc	9540
tctctcgagc	tttcgcagat	ccgggggggg	aatgagatat	gaaaaagcct	gaactcacgg	9600
cgacgtctgt	cgagaagttt	ctgatcgaaa	agttcgacag	cgtctccgac	ctgatgcagc	9660
tctcggaggg	cgaagaattc	cgtgctttca	gcttcgatgt	aggaggggcg	ggatatgtcc	9720
tgccgggtaaa	tagctgcgcc	gatggtttct	acaaagatcg	ttatgtttat	cgggactttg	9780
catcggccgc	gctcccgaat	ccggaagtgc	ttgacattgg	ggagtttagc	gagagcctga	9840
cctattgcat	ctcccgcgt	gcacagggtg	tcacgttgca	agacctgcct	gaaacccaac	9900
tgcccgtgtg	tctacaaccg	gtcgcggagg	ctatgtagtc	gatcgctgcg	gccgatctta	9960
gccagacgag	cgggttcggc	ccattccggac	cgcaagggaat	cgggtcaatac	actacatggc	10020
gtgatttcat	atgcgcgat	gctgatcccc	atgtgtatca	ctggcaaaact	gtgatggagc	10080
acaccgtcag	tgctccgtc	gcgcaggctc	tcgatgatgc	gatgctttgg	gccgaggact	10140
gccccgaagt	ccggcacctc	gtgcacgcgg	atttcggctc	caacaatgtc	ctgacggaca	10200
atggccgcac	aacacgggtc	attgactgga	gcgaggcgat	gttcggggat	tcccaatacg	10260
aggtcgccaa	catctcttcc	tggaggccgt	ggttggcttg	tatggagcag	cagaccgctg	10320
acttcgagcg	gaggcatccg	gagcttgcag	gatcgccacg	actccgggcg	tatatgctcc	10380
gcattgggtc	tgaccaactc	tatcagagct	tggttgacgg	caatttcgat	gatgcagctt	10440
gggcgcaggg	tcgatgcgac	gcaatcgtcc	gatccggagc	cgggactgtc	gggcgtacac	10500

aaatcgcccc	cagaagcgcg	gccgtctgga	ccgatggctg	tgtagaagta	ctcgccgata	10560
gtggaaaccg	acgccccagc	actcgtccga	gggcaaagaa	atagagtaga	tgccgaccgg	10620
atctgtcgat	cgacaagctc	gagtttctcc	ataataatgt	gtgagtagtt	cccagataag	10680
ggaattaggg	ttcctatagg	gtttcgctca	tgtgttgagc	atataagaaa	cccttagtat	10740
gtatttgtat	ttgtaaaata	cttctatcaa	taaaatttct	aattcctaaa	accaaaatcc	10800
agtactaaaa	tccagatccc	ccgaattaat	tcggcggttaa	ttcagatcaa	gcttgacctg	10860
gaatatcgcg	agtaaaactga	aaatcacgga	aaatgagaaa	tacacacttt	aggacgtgaa	10920
atatggcgag	gaaaactgaa	aaagggtgga	aatttagaaa	tgtccactgt	aggacgtgga	10980
atatgggaag	aaaaactgaaa	atcatggaaa	atgagaaaca	tccacttgac	gacttgaaaa	11040
atgacgaaat	cactaaaaaa	cgtgaaaaat	gagaaatgca	cactgaagga	ctccgcggga	11100
attcgattgt	gctagccaat	gtttaacaag	atgtcaagca	caatgaatgt	tggtggttgg	11160
tggtcgtggc	tggtcggtggt	ggaaaattgc	gggtggttca	gcggtagtga	tcggcgatgg	11220
ttggtgtttg	cagcggtgtt	tgatatcgga	atcacttatg	gtggttgtca	caatggaggt	11280
gcgtcatggt	tattggtggt	tggtcatcta	tatatattta	taataatatt	aagtatttta	11340
cctatttttt	acataatttt	tattaaattt	atgcattggt	tgtattttta	aatagttttt	11400
atcgtacttg	ttttataaaa	tattttatta	ttttatgtgt	tatatatta	cttgatgtat	11460
tggaattttt	ctccattggt	ttttctatat	ttataataat	tttcttattt	ttttttgtaa	11520
tattatgtat	tttttcgttt	tataataaat	attttataaa	aaaaatatta	ttttttgtaa	11580
atatatcatt	tacaatgttt	aaaagtcatt	tgtgaatata	ttagctaagt	tgtacttctt	11640
tttgtgcatt	tggtgttgta	catgtctatt	atgattctct	ggccaaaaca	tgtctactcc	11700
tggtcattgg	tttttttttt	ttaagacata	atcattagt	attatatcta	gactgaagga	11760
gggaaacgac	aatctgatca	tgagcggaga	attaagggag	tcacgttatg	acccccgcgg	11820
atgacgcggg	acaagcgtt	ttacgtttgg	aactgacaga	accgcaacgt	tgaaggagcg	11880
actcagccgc	gggtttctgg	agtttaata	gctaagcaca	tacgtcagaa	accattattg	11940
cgcggtcaaa	agtcgcctaa	ggcactatc	agctagcaaa	tatttcttgt	caaaaatgct	12000
ccactgacgt	tccataaatt	cccctcggt	tccaattaga	gtctcatatt	cactctcaat	12060
ccaaataatc	tgacacggat	ctcgagatcg	aattccccgc	gcccgcgaatt	cactagtggg	12120
cccccggtta	cggtcagtc	cttatgttac	gtcctgtaga	aaacccaacc	cgtgaaatca	12180
aaaaactcga	cggtcgtggt	gcattcagtc	tggatcgcca	aaactgtgga	attgagcagc	12240
gttggtggga	aagcgcgtta	caagaaaagg	gggcaattgc	tgtgccaggc	agttttaacg	12300
atcagttcgc	cgatgcagat	attcgtaatt	atgtgggcaa	cgtctggtat	cagcgcgaag	12360
tctttatacc	gaaaggttgg	gcaggccagc	gtatcgtgct	gcgtttcgat	gcggtcactc	12420
atttcgcaa	agtgtgggtc	aataatcagg	aagtgtatga	gcacagggc	gctatagcgc	12480
catttgaagc	cgatgtcacg	ccgtatgtta	ttgccgggaa	aagtgtacgt	atcacagttt	12540
gtgtgaacaa	cgaactgaac	tggcagacta	tcccgcgggg	aatgggtgatt	accgacgaaa	12600
acggcaagaa	aaagcagtc	tacttccatg	tttcttttaa	ctacgcgggg	atccatcgga	12660
gcgtaatgct	ctacaccacg	ccgaacacct	gggtggacga	tatcaccgtg	gtgacgcag	12720
tcgctgcaaga	ctgtaacacc	gcgtctgttg	actggcaggt	gggtggccaat	gggtgatgtc	12780
gcgttgaaat	gcgtgatgcg	gatcaacagg	tgtttgcaac	tggacaaggc	accagcgagg	12840
ctttgcaagt	gggtgaatccg	cacctctggc	aaacgggtga	aggttatctc	tatgaactgt	12900
acgtcacage	caaaagccag	acagagtgtg	atatctaccc	gctgcgcgtc	ggcatccggt	12960
cagtgccgca	gaagggcgaa	cagttcctga	tcaaccacaa	accgttctac	tttatggct	13020
ttggccgtca	tgaagatgcg	gatttgccgc	gcaaaggatt	cgataacgtg	ctgatgggtc	13080
acgatcacgc	attaatggac	tggattgggg	ccaactccta	ccgtacctcg	cattaccctt	13140
acgctgaaga	gatgtcgac	tgggcagatg	aacatggcat	cgtggtgatt	gatgaaactg	13200
cagctgtcgg	ctttaacctc	tctttaggca	ttggtttcga	agcgggcaac	aagccgaaag	13260
aactgtacag	cgaagaggca	gtcaacgggg	aaactcagca	ggcgcactta	caggcgatta	13320
aagagctgat	agcgcgtgac	aaaaaccacc	aaagcgtggt	gatgtggagt	attgccaacg	13380
aaccggatac	ccgtccgcaa	gggtgcacgg	aatatttcgc	gccactggcg	gaagcaacgc	13440
gtaaaactcga	tccgacgcgt	ccgatcacct	gcgtcaatgt	aatgttctgc	gacgctcaca	13500
ccgataccat	cagcgcgtct	tttgatgtgc	tgtgcctgaa	ccgttattac	ggttggtatg	13560
tccaaagcgg	cgatttggaa	acggcagaga	aggtactgga	aaaagaactt	ctggcctggc	13620
aggagaaact	gcacagcccg	attatcatca	ccgaatacgg	cgtggatacg	ttagccgggc	13680
tgactccaat	gtacaccgac	atgtggagtg	aagagtatca	gtgtgcatgg	ctggatatgt	13740
atcacgcggt	ctttgatcgc	gtcagcgccg	tcgctcggtga	acagggtatg	aatttcgcgg	13800
attttgcgac	ctcgcaaggc	atatgtcgcg	ttggcggtaa	caagaagggg	atcttcaccc	13860
gcgacgcgaa	accgaagtgc	gcggcttttc	tgctgcaaaa	acgctggact	ggcatgaact	13920
tcggtgaaaa	accgcagcag	ggaggcaaac	aatgaaatcaa	caactctcct	ggcgaccat	13980
cgctcggtac	agcctcggga	attgctgacc	gagctcgaat	ttccccgatc	gttcaaacat	14040
ttggcaatga	agttttcttaa	gattgaatcc	tgttgcgggt	cttgcatga	ttatcatata	14100
atttctgttg	aattacgtta	agcatgtaat	aattacatgt	taattgcatga	cgttattttat	14160
gagatgggtt	tttatgatta	gagtcgccga	attatacatt	taatacgcga	tagaaaaaaa	14220
aatatagcgc	gcaaaactagg	ataaaattatc	gcgcgcgggt	tcactctatgt	tactagatcg	14280
ggaattcgat	ataaagcttg	gcactggccg	tctgtttaca	acgtcgtgac	tggggaaaaa	14340
ctggcggttac	ccaacttaat	cgccttgcag	cacatcccc	tttcgcccagc	tggtcgtaata	14400
gcgaagaggg	ccgcaccgat	cgcccttccc	aacagttgct	cagcctgaat	ggcgaatgct	14460
agagcagctt	gagcttggat	cagattgtcg	ttccccgcct	tcagttttaa	ctatcagtg	14520

-71-

ttgacaggat atattggcgg gtaaacctaa gagaaaagag cgtttattag aataacggat 14580  
 atttaaaagg gcgtgaaaag gtttatccgt tcgtccattt gtatgtg 14627

<210> 110  
 <211> 9080  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> p18attBZeo (6XHS4) 2eGFP Plasmid

<400> 110  
 cagttgcccgg ccgggtcgcg cagggcgaaac tccccccccc acggetgtct gccgatctcg 60  
 gtcatggccgg gccccgaggc gtcccgggaag ttcgtggaca cgacctccga ccactcggcg 120  
 tacagctcgt ccaggcccgcg caccacacacc caggccaggg tggtgtccgg caccacctgg 180  
 tcctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtgcgtc 240  
 tccacgaagt cccgggagaa cccgagccgg tcgggtccaga actcgaccgc tccggcgacg 300  
 tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360  
 caagtttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatatc gaattcctgc 420  
 agcccccgcg atccgctcac ggggacagcc ccccccaaaa gccccagggg atgtaattac 480  
 gtccctcccc cgctaggggg cagcagcgag ccccccgggg ctccgctccg gtccggcgct 540  
 ccccccgcat ccccgagccg gcagcgtgcg gggacagccc gggcacgggg aaggtggcac 600  
 gggatcgcgt tcctctgaac gcttctcgct gctctttgag cctgcagaca cctgggggat 660  
 acggggcccg ggatccgctc acggggacag ctgctctttg ccccccacca aagccccag ggatgtaatt 720  
 acgtccctcc cccgctaggg ggcagcagcg agccgcccgg ggctccgctc cgggtccggcg 780  
 ctcccccccg atccccgagc cggcagcgtg cggggacagc ccggggcacgg ggaaggtggc 840  
 acgggatcgc acgcttctcg ctgctctttg agcctgcaga cacctggggg 900  
 atacggggcc gcggatccgc tcacggggac agcccccccc caaagcccc agggatgtaa 960  
 ttacgtccct ccccgcttag ggggcagcag cgagccgccc ggggctccgc tccgggtccg 1020  
 cgctcccccc gcctccccga gccggcagcg gcccgggacg ggggaagggt 1080  
 gcacgggatc gctttcctct gaacgcttct cgctgctctt tgagcctgca gacacctggg 1140  
 ggatacgggg ccgcggatcc gctcacgggg acagccccc cccaaagccc ccagggatgt 1200  
 aattacgtcc tcccccgctc agggggcagc agcagccgc cgggggctcc gctccggctc 1260  
 ggcgctcccc ccgcatcccc ctgctttcct ctcgctggga cagcccgggg acggggaagg 1320  
 tggcacggga tcgctttcct ctgaacgctt ctgcgtgctc tttgagcctg cagacacctg 1380  
 ggggatacgg ggccgcgat ccgctcacg ctgagccccc ggacagcccc ccccagggat 1440  
 gtaattacgt cccctccccg ctagggggca gcagcagacc gcccggggct cccggtccgg 1500  
 ccggcgctcc ccccgcatcc ccgagccggc agcgtgcggg gacagcccgg gcacggggaa 1560  
 ggtggcacgg gatcgctttc ctctgaacgc tctcgctgc tctttgagcc tgcagacacc 1620  
 tgggggatac ggggcccggg atccgctcac ggggacagcc ccccccaaaa gccccagggg 1680  
 atgtaattac gtccctcccc cgctaggggg cagcagcgag ccgcccgggg ctccgctccg 1740  
 gtccggcgct cccccgcgt ccccgagccg gcagcgtgcg gggacagccc gggcagcggg 1800  
 aaggtggcac gggatcgctt tcctctgaac gcttctcgct gctctttgag cctgcagaca 1860  
 cctgggggat acggggcggg ggatccacta gttattaata gtaatcaatt acgggggtcat 1920  
 tagttcatag cccatatatg gagttccgcy ttacataact tacggtaaat ggcccgctcg 1980  
 gctgacggcc caacgacccc cgcccattga cgtcaataat gacgtatgtt cccatagtaa 2040  
 cgccaatagg gactttccat tgacgtcaat ggggtgacta tttacggtaa actgccact 2100  
 tggcagtaca tcaagtgtat catatgccaa gtacgcccc tattgacgtc aatgacggt 2160  
 aatggcccg ctaggattat gccagtaga tgaccttatg ggactttcct acttggcagt 2220  
 acatctacgt attagtcac tcctctgccc gctattacca tgggtcgagg tgagccccac gttctgcttc 2280  
 actctcccca tctccccccc ctcccccccc ccaattttgt atttatttat tttttaatta 2340  
 ttttgtgcag cgatgggggg gggggggggg ggggcccggg ccaggcgggg cggggcgggg 2400  
 cgagggggcg ggcggggcga ggcggagagg tgccggcgga gccaatcaga gcggcgcgct 2460  
 ccgaaagtgt ccttttatgg cgaggccggg gcggcgggcg ccctataaaa agcgaagcgc 2520  
 gcggcgggcg ggagtgcgt cggtgccttc gcccctgcc cgccctcgcg ccgctcgcg 2580  
 ccgcccggcc cggtctgac tgaccgcgtt actccacag gtgagcgggc gggacggccc 2640  
 ttctctcccg ggctgtaatt agcgtttggt ttaatgacgg ctggtttcct ttctgtggct 2700  
 cgctgaaagc cttaaagggc tccgggaggg cctttgtgc ggggggggagc ggctcgggg 2760  
 gtgcgtgcgt gtgtgtgtgc gtggggagcg ccgctgcggg cccgcgctgc cggcgcgctg 2820  
 tgagcgtgc gggcgcgggc gggggctttg tgcgtccgc gtgtgcgcga ggggagcgcg 2880  
 gccggggcg gtgccccg cggtcgaggg gggcgaggg gaacaaaggc tgcgtgcggg 2940  
 gtgtgtgcgt gggggggtga gcagggggtg tggcgcgggc ggtcgggctg taaccccccc 3000  
 ctgcaccccc ctccccgagt tgctgagcac gggccggctt cgggtgcggg gctccgtgcg 3060  
 gggcgctcgc cgggctcgc cgtgccgggc gggggagggc ggcaggtggg ggtgcgggg 3120  
 gggggcgggc cgccctcggg ccgggagcgg tttatggtaa ccccgagcgc 3180  
 ccggcggtcg tcgaggcgcg gccagccgca tctatggtaa tcgtgcgaga 3240  
 gggcgcgagg acttcctttg tcccaaactt ggcggagccg aaatctggga ggcgcgcgcg 3300

cacccccctct	agcggggcgcg	ggcgaagcgg	tgcgggcgccc	gcaggaagga	aatggggcggg	3360
gagggccttc	gtgcgtcgcc	gcgcgcgcgt	ccccctctcc	atctccagcc	tgcgggctgc	3420
cgagggggga	cggctgcctt	cgggggggac	ggggcagggc	gggggttcggc	ttctggcggtg	3480
tgaccggcg	ctctagagcc	tctgctaacc	atgttcatgc	cttcttcttt	ttcctacagc	3540
tccctgggcaa	cgtgctggtt	gttgtgctgt	ctcatcattt	tggcaaaagaa	ttcgccacca	3600
tggtagacaa	gggcgaggag	ctgttcaccg	gggtggtgcc	catcctgggtc	gagctggacg	3660
gcgacgtaaa	cggccacaag	ttcagcgtgt	ccggcgaggg	cgagggcgat	gccacctacg	3720
gcaagctgac	cctgaagttc	atctgcacca	ccggcaagct	gccccgtgcc	tggcccaccc	3780
tcgtagccac	cctgacctac	ggcgtgcagt	gcttcagccg	ctaccccgac	cacatgaagc	3840
agcacgactt	cttcaagttc	gccatgcccc	aaggctacgt	ccaggagcgc	accatcttct	3900
tcaaggacga	cggcaactac	aagaccgcgc	ccgaggtgaa	gttcgagggc	gacacctggg	3960
tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	cggcaacatc	ctggggcaca	4020
agctggagta	caactacaac	agccacaacg	tctatatcat	ggccgacaag	cagaagaacg	4080
gcatcaaggt	gaacttcaag	atccgccaca	acatcgagga	cggcagcgtg	cagctcgccg	4140
accactacca	gcagaacacc	cccacgcggc	acggccccgt	gctgctgccc	gacaaccact	4200
acctgagcac	ccagtccgcc	ctgagcaaa	accccaacga	gaagcgcgat	cacatggtcc	4260
tgctggagtt	cgtgaccgcc	gccgggatca	ctctcggtcat	ggacgagctg	tacaagtaag	4320
aattcactcc	tcagggtgcag	gctgcctatc	agaaggtggt	ggctggtgtg	gccaatgccc	4380
tggctcacia	ataccactga	gatctttttc	cctctgccaa	aaattatggg	gacatcatga	4440
agccccctga	gcactctgact	tctggctaat	aaaggaaatt	tatttttcat	gcaatagtgt	4500
gttggaaatt	tttgtgtctc	tcactcgga	ggagcatatgg	gagggcaaat	catttaaac	4560
atcagaatga	gtatttggtt	tagagttagg	caacatagtc	catatgctgg	ctgccatgaa	4620
caaaggtggc	tataaagagg	tcatcagtat	atgaaacagc	cccctgctgt	ccattcctta	4680
ttccatagaa	aagccttgac	ttgaggttag	atttttttta	tattttgttt	tgtgttattt	4740
ttttctttaa	catccctaaa	attttcctta	catgttttac	tagccagatt	tttctctctc	4800
tccctgactac	tcccagtcac	agctgtccct	cttctcttat	gaagatccct	cgacctgcag	4860
cccaagcttg	catgcctgca	ggtcgactct	agtggtatcc	cgcgcccgta	tccccagggt	4920
gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	agcgatcccc	tgccaccttc	4980
cccgtgcccc	ggctgtcccc	gcacgctgcc	ggctcgggga	tgccgggggga	gcgcgggacc	5040
ggagcggagc	cccgggcggc	tcgctgctgc	cccctagcgg	gggagggacg	taattacatc	5100
cctggggggt	ttgggggggg	gctgtccccg	tgagcggatc	cgccggcccc	tatccccag	5160
gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	aaagcgatcc	cgtgccacct	5220
tccccgtgcc	cggcgtgtcc	ccgcacgctg	ccggctcggg	gatgcggggg	gagcgccgga	5280
ccggagcggg	gccccggggc	gctcgctgct	gccccctagc	gggggagggga	cgtaattaca	5340
tccccggggg	ctttgggggg	gggctgtccc	cgtgagcggg	tccgcggccc	cgtatcccc	5400
aggtgtctgc	aggctcaaag	agcagcgaga	agcgttcaga	ggaaagcgat	cccgtgccac	5460
cttccccgtg	ccccggcgtg	ccccgcacgc	tgccggctcg	gggatgcggg	gggagcgccg	5520
gaccggagcg	gagccccggg	cggctcgctg	ctgcccccta	gcgggggagg	gacgtaatta	5580
catccctggg	ggctttgggg	gggggctgtc	cccgtgagcg	gatccgcggc	ccgctatccc	5640
ccaggtgtct	gcaggctcaa	agagcagcga	gaagcgttca	gaggaagcgc	atcccgctgc	5700
acettccccg	tgccccgggt	gtccccgcac	gctgccggct	cggggatgcg	gggggagcgc	5760
cggaccggag	cggagccccg	ggcggctcgc	tgctgcccc	tagcggggga	gggacgtaat	5820
tacatccctg	ggggggggct	ggggggggct	tccccgtgag	cggatccgcg	gccccgtatc	5880
ccccaggtgt	ctgcaggctc	aaagagcagc	gagaagcgtt	cagaggaaag	cgatcccgtg	5940
ccaccttccc	cgtgccccgg	ctgtccccgc	acgtgccccg	ctcggggatg	cgggggggagc	6000
gccggaccgg	agcggagccc	cgggcggctc	gctgctgccc	cctagcgggg	gagggacgta	6060
attacatccc	tgggggcttt	gggggggggc	tgtccccgtg	agcggatccg	cggccccgta	6120
tccccaggtt	gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	agcgatcccc	6180
tgccaccttc	cccgtgcccc	ggctgtcccc	gcacgctgcc	ggctcgggga	tgccggggga	6240
gcgcgggacc	ggagcggagc	cccgggcggc	tcgctgctgc	cccctagcgg	gggagggacg	6300
taattacatc	cctggggggt	ttgggggggg	gctgtccccg	tgagcggatc	cgccgggctg	6360
caggaattcg	taatcatggt	catagctgtt	tcctgtgtga	aattgttatc	cgctcacaat	6420
tccacacaac	atacagcccg	gaagcataaa	gtgtaaagcc	tggggtgcct	aatgagtgag	6480
ctaactcaca	ttaattgcgt	tgctgtcact	gcccgttttc	cagtcgggaa	acctgtcgtg	6540
ccagctgcat	taatgaatcg	gccaacgcgc	ggggagaggc	ggtttgcgta	ttgggcgctc	6600
ttccgcttcc	tcgctcactg	actcgctgcg	ctcggtcgtt	cggctgcggc	gagcgggtatc	6660
agctcactca	aaggcggtaa	tacggttatc	cacagaatca	ggggataaac	caggaaagaa	6720
catgtgagca	aaaaggccagc	gaaccgtaaa	aaggcccgct	aaggcccgct	tgtgaaagct	6780
tttccatagg	ctccgcccc	ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagaggtg	6840
gcgaaccggc	acaggactat	aaagatacca	ggcgtttccc	cctggaagct	ccctcgtgcg	6900
ctctcctgtt	ccgacctg	cgcttacggg	ataccgttcc	gcctttctcc	cttcgggaag	6960
cgtggcgctt	tctcatagct	cacgctgtag	gtatctcagt	tcggtgtagg	tcggtcgctc	7020
caagctgggc	tgtgtgcacg	aacccccgt	tcagcccgac	cgctgcgcct	tatccggtaa	7080
ctatcgtctt	gagtaagaca	cggtatgtag	ccactggcag	gagttcttga	cagccactgg	7140
taacaggatt	agcagagcga	ggatgttagg	cggtgctaca	gctctgctga	agtgggtggc	7200
taactacggc	tacactagaa	ggacagtatt	tggatatctgc	accaccgctg	agccagttac	7260
cttcggaaaa	agagttggta	gctcttgatc	cggcaaacaa		gtagcgggtg	7320



-73-

tttttttgggt	tgcaagcagc	agattacgcg	cagaaaaaaa	ggatctcaag	aagatccttt	7380
gatcttttct	acgggggtctg	acgctcagtg	gaacgaaaaac	tcacgtttaag	ggatttttgggt	7440
catgagattta	tcaaaaaggga	tcttcaccta	gatccttttta	aattaaaaaat	gaagtttttaa	7500
atcaatcttaa	agtatatatg	agtaaaacttg	gtctgacagt	taccaatgct	taatcagtgga	7560
ggcacctatc	tcagcgatct	gtctatttcg	ttcatccata	gttgcctgac	tccccgtcgt	7620
gtagataaact	acgatacggg	agggcttacc	atctggcccc	agtgcctgcaa	tgataccgcg	7680
agaccacacgc	tcaccggctc	cagattttatc	agcaataaac	cagccagccg	gaagggccga	7740
gcgcagaagt	ggctctgcaa	ctttatccgc	ctccatccag	tctattaatt	gttgccggga	7800
agctagattcg	agtagttcgc	cagttaatag	tttgcgcaac	gttggttgcca	ttgctacagg	7860
catcggtggtg	tcacgctcgt	cgtttggtat	ggcttcattc	agctccggtt	cccaacgatc	7920
aaggcgagtt	acatgatccc	ccatgttgtg	caaaaaagcg	gttagctcct	tccgtcctcc	7980
gatcgttggtc	agaagtaagt	tggccgcagt	gttatcactc	atgggttatgg	cagcactgca	8040
taattctctt	actgtcatgc	catccgtaag	atgcttttct	gtgactgggtg	agtactcaac	8100
caagtcattc	tgagaatagt	gtatgcggcg	accgagttgc	tcttgcccgg	cgtcaataacg	8160
ggataaatacc	gcgccacata	gcagaacttt	aaaagtgtc	atcattggaa	aacgttcttc	8220
ggggcgaaaa	ctctcaagga	tcttaccgct	gttgagatcc	agttcgatgt	aaccactcgc	8280
tgaccccaac	tgatcttcag	catcttttac	tttcaccagc	gtttctgggt	gagcaaaaaac	8340
aggaaggcaa	aatgccgcaa	aaaagggaat	aaggcgaca	cggaaatgtt	gaatactcat	8400
actcttcctt	tttcaatatt	attgaagcat	ttatcagggt	tattgtctca	tgagcgggata	8460
catatttgaa	tgtatttaga	aaaataaaca	aataaggggt	ccgcgcacat	ttccccgaaa	8520
agtgcacct	gacgtagtta	acaaaaaaaa	gcccgcgaa	gcggttcta	ttaccaagcg	8580
aagcgccatt	cgccattcag	gctgcgcaac	tgttggaag	ggcgatcggt	gcggtcctct	8640
tcgctattac	gccagctggc	gaaaggggga	tgtgctgcaa	ggcgattaa	ttgggttaacg	8700
ccagggtttt	cccagtcacg	acgttgtaaa	acgacggcca	gtccgtaata	cgactcactt	8760
aaggccttga	ctagaggggtc	gacggtatac	agacatgata	agatacattg	atgagtttgg	8820
acaaaccaca	actagaatgc	agtgaaaaaa	atgctttatt	tgtgaaattt	gtgatgctat	8880
tgctttattt	taaaccttta	taagctgcaa	taacaagtt	gggttggtcg	aagaactcca	8940
gcatgagatc	cccgcgctgg	aggatcatcc	agccggcgtc	ccggaaaacg	attccgaagc	9000
ccaacctttc	atagaaggcg	gcggtggaat	cgaatctcg	tagcacgtgt	cagtcctgct	9060
cctcgccac	gaagtgcacg					9080

&lt;210&gt; 111

&lt;211&gt; 4223

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pLIT38attBBSRpolyA10 Plasmid

&lt;400&gt; 111

gttaactacg	tcaggtggca	cttttcgggg	aaatgtgcgc	ggaaccacct	tttgtttatt	60
tttctaataa	cattcaaata	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	120
ataatattga	aaaagggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	180
ttttgcggca	ttttgccttc	ctgttttttg	tcacccagaa	acgctgggtga	aagtaaaaga	240
tgctgaagat	cagttgggtg	cacgagtggtg	ttacatcgaa	ctggatctca	acagcgggta	300
gatccttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagttct	360
gctatgtggc	gcggtattat	cccgtgttga	cgccgggcaa	gagcaactcg	gtcgccgcct	420
acactattct	cagaatgact	tgggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	480
tggcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	atgagtgata	acactgcggc	540
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaacaacat	600
gggggatcat	gtaactcgcc	ttgatcggtg	ggaaccggag	ctgaatgaag	ccataccaaa	660
cgacgagcgt	gacaccacga	tgccgtgtagc	aatggcaaca	acgttgcgca	aactatttaac	720
tggcgaaact	cttactctag	cttcccggca	acaattaata	gactggatgg	aggcgggata	780
agttgcagga	ccacttctgc	gctcggtcct	tccggctggc	tggtttattg	ctgataaaatc	840
tggagccggt	gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	atggttaagcc	900
ctcccgatc	gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	960
acagatcgct	gagatagggtg	cctcactgat	taagcattgg	taactgtcag	accaagttta	1020
ctcatatata	ctttagattg	atttaacccc	gttgataatc	agaaaaagccc	caaaaacagg	1080
aagattgtat	aagcaaatat	ttaaattgta	aacgttaata	ttttgttaaa	attcgcgtta	1140
aatttttgtt	aaatcagctc	attttttaac	caataggccg	aaatcggcaa	aatcccttat	1200
aatcaaaaag	aatagcccga	gatagggttg	agtgttgttc	cagtttgga	caagagttcca	1260
ctattaaaga	acgtggactc	caacgtcaaa	gggcgaaaaa	ccgtctatca	gggcgatggc	1320
ccactacgtg	aaccatcacc	caaatcaagt	tttttggggg	cgaggtgccg	taaagcacta	1380
aatcggaacc	ctaaaggag	ccccgattt	agagcttgac	ggggaaaagc	aacgtggcga	1440
gaaaggaagg	gaagaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcgggtca	1500
cgctgcgcgt	aaccaccaca	ccgcgcgcgc	ttaatgcgcc	gctacagggc	gcgtaaaagg	1560
atctaggtga	agatcctttt	tgataatctc	atgacaaaaa	tcccttaacg	tgagttttcg	1620

-74-

```
ttccactgag cgtcagaccc cgtagaaaag atcaaaggat cttcttgaga tccttttttt 1680
ctgcgcgtaa tctgtgctt gcaaacaaaa aaaccaccgc taccagcggg ggtttgtttg 1740
ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag agcgagata 1800
ccaaatactg ttcttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca 1860
ccgcctacat acctcgctct gctaatacctg ttaccagtgg ctgctgccag tggcgataag 1920
tcgtgtctta ccgggttggg ctcaagacga tagttaccgg ataaggcgca gcggctgggc 1980
tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 2040
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aaggagagaaa ggcggacagg 2100
tatccggtaa cgggcagggt cggaacagga gagcgcacga gggagcttcc ggggagcttcc 2160
gcctgggtatc tttatagtcg tgctcggttt cgccacctct gacttgagcg gacttgagcg 2220
tgatgctcgt cagggggggcg gaggcctatgg aaaaacgcca gcaacgcggc ctttttacgg 2280
ttcctggcct tttgctggcc ttttgcctac agtttagctca agtttagctca ctctattaggc 2340
accccgaggct ttacacttta gctatgacca tgctatgttg tgtggaattg tgagcggata 2400
acaatttcac acaggaaaca cgtgcaattg ggcgccaagc aagctacgta agctacgta 2460
ctagtggggc tactaacttg agaattagta gcaattcgta ttctctgcag gattgaagcc 2520
tgctttttta tactaacttg gaagttagcga cgagtaactg ggtaccacat aacatttctc 2580
aacaagatct agaattagta gcaattcgta cgagtaactg ggtaccacat aacatttctc 2640
ataaacatca tgtgggagcg gcaattcgta cgagtaactg ggtaccacat aacatttctc 2700
atattgaagc gtatatagga gcaattcgta cgagtaactg ggtaccacat aacatttctc 2760
cagtttcgaa tggacaaaag gattttgaca gtccttgtgg tagaaatgaa tggcaagtta 2820
acgaagttag accagattgt actcattcca cctgaggctg cccgaaatta aaagttttac 2880
cagactatgc cagattgaaga ttgctgctg gcatccagga aaccagcagc cgggagctgc cccgagctgc 2940
cgaattgaaga ttgctgctg gcatccagga aaccagcagc cgggagctgc cccgagctgc 3000
cgaattgaaga ttgctgctg gcatccagga aaccagcagc cgggagctgc cccgagctgc 3060
cataatttga ccaactacct tggtataatg gaactgatga aagaaatgcc atctagtgat 3120
gtgtataatg gaactgatga aagaaatgcc atctagtgat aaaggtagaa gtttagtaat 3180
aagaaatgcc atctagtgat aaaggtagaa gtttagtaat gtttagtaat 3240
aaaagaagag gtttagtaat gtttagtaat gtttagtaat gtttagtaat 3300
aagaaatgcc atctagtgat aaaggtagaa gtttagtaat gtttagtaat 3360
ataacagtta taatcataac ctatgtctca ataggaataa tttgatgtat 3420
ctatataata ctatgtctca ataggaataa tttgatgtat tacttgcttt 3480
ataacagtta taatcataac ctatgtctca ataggaataa tttgatgtat 3540
aagaaatgcc atctagtgat aaaggtagaa gtttagtaat gtttagtaat 3600
aatagatcaa caaatttcac gtctccggat gtacaggcat gtccttttac 3660
gtctccggat gtacaggcat gtccttttac cgccttgcca gcacatcccc 3720
cgccttgcca gcacatcccc caacagttgc cccgcttcgg cgggcttttt 4200
cccgcttcgg cgggcttttt ttt
```

&lt;210&gt; 112

&lt;211&gt; 5855

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pCX-LamIntR Plasmid

&lt;400&gt; 112

```
gtcgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60
gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgccg ggctgaccgc 120
ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180
ggactttcca ttgagtcaca tgggtggact attacggta aactgcccac ttggcagtag 240
atcaagtgtg tcatatgcca agtacgcccc ctattgacgt caatgacggg aaatggcccc 300
cctggcatta tgcccagtag atgaccttat gggactttcc tacttggcag tacatctacg 360
tattagtcag cgtattacc atgggtcgag gtgagcccca cgttctgctt cactctcccc 420
atctccccc cctcccccac cccaattttg tattttatta ttttttaatt attttgtgca 480
gcgatggggg cggggggggg gggggcgccg gccaggcggg gcggggcggg gcgagggcg 540
ggcgggggcg agcgggagag gtgcggcgag gcccaatcag agcgggcgcg tccgaaagtt 600
tccttttatg gcgagggcgg ggcggcgggc gccctataaa aagcgaagcg cgcggcgggc 660
gggagtcgct gcgttgcttc cgccttcgga cccgctccgc gccgctccgc gccgcccggc 720
ccggtcttga ctgaccgcgt tactccaca ggtgagcggg cgggacggcc cttctcctcc 780
```



gggctgtaaat	tagcgcttgg	tttaatgaag	gctcgtttct	tttctgtggc	tgcgtgaaag	840
ccttaaaggg	ctccgggagg	gccctttgtg	cgggggggag	cggtccgggg	ggtgcgtgcg	900
tgtgtgtgtg	cgtggggagc	gccgcgtgcg	gcccgcgctg	cccgccggct	gtgagcgctg	960
cgggcgcgcc	gcggggcttt	gtgcgctccg	cgtgtgctcg	aggggagcgc	ggccggggggc	1020
ggtgccccgc	ggtgccccgg	ggctgcgagg	ggaacaaagg	ctgcgtgctg	ggtgtgtgcg	1080
tgggggggtg	agcagggggg	gtggggcgcg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cgggccggct	tcgggtgctg	ggctccgtgc	ggggcggtggc	1200
gcggggctcg	ccgtgcccgg	cggggggtgg	cggcaggctg	gggtgccggg	cgggggcgggg	1260
ccgcctcggt	ccggggaggg	ctcgggggag	gggcgcggcg	gccccggagc	gccggcggtct	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttatggta	atcgctgcgag	agggcgcgagg	1380
gacttccttt	gtcccaaata	tggcgaggcg	gaaatctggg	aggcgccgcc	gcacccccctc	1440
tagcgggcgc	ggcggaagcg	gtgcggcgcc	ggcaggaagg	aaatggcgcg	ggagggcctt	1500
cgtgcgtcgc	cgcgcgcgcg	tccccttctc	catctccagc	ctcggggctg	ccgcagggggg	1560
acggctgcct	tcggggggga	cggggcaggg	cggggttcgg	cttctggcgt	gtgacgggca	1620
gctctagagc	ctctgctaac	catgttcatg	ccttcttctt	tttctacag	ctcctgggca	1680
acgtgctggt	tgttgtgctg	tctcatcatt	ttggcaaaag	attcatggga	agaaggcgaa	1740
gtcatgagcg	ccgggattta	ccccctaacc	tttatataag	aaacaatgga	tattactgct	1800
acagggaccc	aaaggacggg	aaagagtttg	gattaggcag	agacaggcga	atcgcaatca	1860
ctgaagctat	acaggccaac	attgagttat	tttcaggaca	caaacacaaag	cctctgacag	1920
cgagaatcaa	cagtgataat	tccgttacgt	tacattcatg	gcttgatcgc	tacgaaaaaa	1980
tcccgggcgc	ggcggaatc	aagcagaaga	cactcataaa	ttacatgagc	aaaattaaag	2040
caataaggag	gggtctgcct	gatgctccac	ttgaagacat	caccacaaaa	gaaattgcgg	2100
caatgctcaa	tggtatacata	gacgagggca	aggcggcgtc	agccaagtta	atcagatcaa	2160
cactgagcga	tgcattccga	gaggcaatag	ctgaaggcca	tataacaaca	aaccatgtcg	2220
ctgccactcg	cgcagcaaaa	tctagagtaa	ggagatcaag	acttacggct	gacgaatacc	2280
tgaataattta	tcaagcagca	gaatcatcac	catgttggct	cagacttgca	atggaactgg	2340
ctgttgtttac	cgggcaacga	gttgggtgatt	tatgcgaaat	gaagtggctc	gatatcgtag	2400
atggatatct	ttatgtcgag	caaagcaaaa	caggcgtaaa	aattgccatc	ccaacagcat	2460
tgcataattga	tgtctccgga	atatcaatga	aggaaacact	tgataaatgc	aaagagattc	2520
ttggcgaggc	aaccataatt	gcacttaact	gtcgcgcaacc	gctttcatcc	ggcacagtat	2580
caaggatatt	tatgcgcgca	cgaaaagcat	caggctcttc	cttcgaaggg	gatccgccta	2640
cctttcacga	gttgcgcagc	ttgtctgcaa	gactctatga	gaagcagata	agcgataagt	2700
tgtctcaaca	cttctccggg	cataagtcgg	acaccatggc	atcacagtat	cgtgatgaca	2760
gaggcaggga	gtgggacaaa	attgaaatca	aataagaatt	cactcctcag	gtgcaggctg	2820
cctatcagaa	gggtgggtgg	gggtgtggcca	atgccttgcc	tcacaaatac	cactgagatc	2880
ttttccctc	tgccaaaaat	tatggggaca	tcatgaaacc	ccttgagcat	ctgacttctg	2940
gctaataaag	gaaatttatt	ttcattgcaa	tagtgtgttg	gaattttttg	tgtctctcac	3000
tcggaaggac	atatggggagg	gcaaatcatt	taaaacatca	gaatgagtat	ttggttttaga	3060
gtttggcaac	atatggccata	tgctggctgc	catgcaacaa	gggtggctata	aagaggtcat	3120
cagtatatga	aacagccccc	tgctgtccat	tccttattcc	atagaaaagc	cttgacttga	3180
ggttagatttt	tttttatatt	ttgttttctg	ttattttttt	ctttaacatc	cctaaaatttt	3240
tccttcatatg	ttttactagc	cagatttttc	ctctctctct	gactactccc	agtcacttctg	3300
gtccctcttc	tcttatgaag	atccctcgac	ctgcagccca	agcttgccgt	aatcatggctc	3360
atagctgttt	cctgtgtgaa	attgttatcc	gtccacaatt	ccacacacaa	tacgagccgg	3420
aagcataaag	tgtaaaagcct	gggtgacctc	atgagtgcgc	taactcacat	taattgcgtt	3480
gcgctcactg	cccgttttcc	agtccggaaa	cctgtcgtgc	cagcgggatcc	gcactctcaat	3540
tagtcagcaa	ccatagtcct	gcccctaact	ccgcccatac	cgccccctaac	tcgcgccagt	3600
tccgcccatt	ctccgccccca	tggctgacta	atttttttta	tttatgcaga	ggccgaggcc	3660
gctcgggct	ctgagctatt	ccagaagtag	tgaggaggct	tttttggagg	cctaggctttt	3720
tgcaaaaagc	taacttggtt	attgtagctt	ataatgggta	caaataaagc	aatagcatca	3780
caaatttcac	aaataaaagca	tttttttcac	tgcatcttag	ttgtgggtttg	tccaaactca	3840
tcaatgtatc	ttatcatgtc	tggatccgct	gcattaatga	atcgggccaac	gcgcgggggag	3900
aggcggtttg	cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	3960
cgttcggtcg	cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	4020
atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaagg	cagcaaaaagg	ccaggaaccg	4080
taaaaaggcc	gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacaa	4140
aatcgacgc	tcaagtccga	ggtggcgaaa	cccgacagga	ctataaagat	accaggcggt	4200
tccccctgga	agctccctcg	tgcgctctcc	tggtccgacc	ctgcccgtta	ccggatccct	4260
gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcaa	tgctcacgct	gtaggtatct	4320
cagttcgggtg	taggtcgttc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	4380
cgaccgctgc	gccttatccg	gtaactatcg	tcttgagtc	aaaccggtaa	gacacgactt	4440
atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	gcgaggatg	tagggcggtgc	4500
tacagagttc	ttgaagtggg	ggcctaacta	cggtacact	agaaggacag	tatttggtat	4560
ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaaaggtt	ggtagctctt	gatccgggcaa	4620
acaaaccacc	gctggtagcg	gtgggtttttt	tggttgcaag	cagcagatta	cgcgagaaa	4680
aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaaacga	4740
aaactcacgt	taagggtattt	tggctcatgag	attatcaaaa	aggatcttca	cctagatcct	4800

-76-

tttaaatttaa	aatgaagtt	ttaaataaat	ctaaagtata	tatgagtaaa	cttgggtctga	4860
cagttaccaa	tgcttaata	gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcatc	4920
catagttgcc	tgactccccg	tcgtgtagat	aactacgata	cgggagggct	taccatctgg	4980
ccccagtgct	gcaatgatac	cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	5040
aaaccagcca	gccggaagg	cgcgagcgcg	aagtggctct	gcaactttat	ccgctccat	5100
ccagttctatt	aattgttgcc	gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	5160
caacgttggt	gccattgcta	caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	5220
attcagctcc	ggttcccaac	gatcaaggcg	agttacatga	tccccatgt	tgtgcaaaaa	5280
agcggttagc	tccttcggtc	ctccgatcgt	tgtcagaagt	aagttggccg	cagtggtatc	5340
actcatgggt	atggcagcac	tgcataattc	tcttactgtc	atgccatccg	taagatgctt	5400
ttctgtgact	ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	5460
ttgctcttgc	ccggcgctcaa	tacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	5520
gctcatcatt	ggaaaaacgtt	cttcggggcg	aaaactctca	aggatcttac	cgctggttag	5580
atccagttcg	atgtaaccca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	5640
cagcgtttct	gggtgagcaa	aaacaggaag	gcaaaaaagg	gaataaaggc	gaataaaggc	5700
gacacggaag	tggtgaatac	tcataactctt	cctttttcaa	tattattgaa	gcatttatca	5760
gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	5820
ggttcgcgcg	acatttcccc	gaaaagtgcc	acctg			5855

&lt;210&gt; 113

&lt;211&gt; 4346

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pSV40-193AttpsensePur Plasmid

&lt;400&gt; 113

ccggtgcccc	caccatcccc	tgaccacgcg	ccctgacccc	tcacaaggag	acgaccttcc	60
atgaccgagt	acaagcccac	ggtgcgcctc	gccaccgcgc	acgacgtccc	ccgggccgta	120
cgaccctcgc	ccgcgcgctt	cgccgactac	cccgccacgc	gccacaccgt	cgaccgggac	180
cgccacatcg	agcgggtcac	cgagctgcaa	gaactcttcc	tcacgcgcgt	cgggctcgac	240
atcggaagg	tgtgggtcgc	ggacgacggc	gcgcggtggg	cggtctggac	cacgccggag	300
agcgtcgaag	cgggggcggt	gttcgcccag	atcgccccgc	gcatggccga	ggtgagcggt	360
tcccggctgg	ccgcgcgagc	acagatggaa	ggcctcctgg	cgccgcaccg	gcccgaaggag	420
cccgctgggt	tcctggccac	cgctggcgct	tcgccccacc	accagggcaa	gggtctgggc	480
agcgcgctcg	tgctcccccg	agtggaggcg	gccgagcgcg	ccgggggtgcc	cgcttccctg	540
gagacctccg	cgccccgcaa	cctccccctt	tacgagcggc	tcggcttcac	cgtcaccgcc	600
gacgtcgagg	tgcccgaagg	accgcgcacc	tgggtcatga	cccgaagcc	cggtcgctga	660
cgcccccccc	acgacccgca	gcgccccgac	gaaaggagcg	cacgacccca	tggtcccgac	720
cgaagccgac	ccggggcgcc	ccgcgcgacc	cgcacccgcg	cccaggcccc	accgactcta	780
gaggatcata	atcagccata	ccacatttgt	agaggtttta	cttgctttaa	aaaacctccc	840
acacctcccc	ctgaacctga	aacataaaat	gaatgcaatt	gttgttgtta	acttggttat	900
tgcagcttat	aattggttaca	aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	960
tttttctactg	cattctagtt	gtggtttgtc	caaactcatc	aatgtatctt	atcatgtctg	1020
gatccgcgcc	ggatccctta	ttaagtctag	agtgcactgt	ttaaacctgc	aggcatgcaa	1080
gcttggcgta	atcatgggtca	tagctgtttc	ctgtgtgaaa	ttgttatccg	ctcacaaattc	1140
cacacaacat	acgagccgga	agcataaagt	gtaaaagcctg	gggtgcctaa	tgagttagct	1200
aactcacatt	aattgcgttg	cgctcactgc	ccgctttcca	gtcgggaaac	ctgtcgtgcc	1260
agctgcatta	atgaatcggc	caacgcgcgg	ggagaggcgg	tttgcgtatt	gggcgctctt	1320
ccgcttcttc	gctcactgac	tcgctgcgct	cggtcgttcg	gctgcggcga	gcggtatcag	1380
ctcactcaaa	ggcggttaata	cggttatcca	cagaatcagg	ggataacgca	ggaaagaaca	1440
tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	ggccgcgttg	ctggcggtttt	1500
tccataggct	ccgcccccc	gacgagcatc	acaaaaatcg	acgctcaagt	cagagggtggc	1560
gaaacccgac	aggactataa	agataccagg	cgtttcccc	tggaagctcc	ctcgtgcgct	1620
ctcctgttcc	gaccctgccc	cttaccggat	acctgtccgc	ctttctccct	tcgggaagcg	1680
tggcgctttc	ctatagctca	cgctgtaggt	atctcagttc	ggtgtaggtc	ggtcgctcca	1740
agctgggctg	tgtcacgaa	ccccccgttc	agccccagcg	ctgcgcctta	tccggtaact	1800
atcgtcttga	gtccaacccg	gtaagacacg	acttatcgcc	actggcagca	gccactggta	1860
acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	gttcttgaag	tggtggccta	1920
actacggcta	cactagaagg	acagtatttg	tctgctgcgc	tctgctgaag	ccagttaact	1980
tcggaaaaag	agttggtagc	tcttgatccg	gcaaaacaa	caccgctggt	agcgggtggt	2040
tttttgtttg	caagcagcag	attacgcgca	gaaaaaaaag	atctcaagaa	gatcccttga	2100
tcttttctac	ggggtctgac	gctcagtgga	acgaaaaactc	acgttaaggg	atttttggta	2160
tgagattatc	aaaaaggatc	ttcacctaga	tcctttttaa	ttaaaaatga	agtttttaaat	2220
caatctaaag	tatatatgag	taaacttggt	ctgacagtta	ccaatgctta	atcagtgagg	2280
cacctatctc	agcgatctgt	ctatttcggt	catccatagt	tgccctgactc	cccgctcggt	2340

-77-

```

agataactac gatacggggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag 2400
acccacgctc accgggctcca gatttatcag caataaacca gccagccgga agggccgagc 2460
gcagaagtgg tccctgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag 2520
ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tggtgccatt gctacaggca 2580
tcgtgggtgc acgctcgctg tttgggtatg cttcattcag ctccggttcc caacgatcaa 2640
ggcgagttac atgatcccc atgttggtgca aaaaagcggg tagctccttc ggtcctccga 2700
tcgttgtagc aagtaagttg gccgcagtgat tatcactcat gggtatggca gcaactgcata 2760
attctcttca tgtcatgcca tccgtaagat gcttttctgt gactgggtgag tactcaacca 2820
agtcattctg agaatagtgt atgcggcgac cgagttgctc ttgcccggcg tcaatacggg 2880
ataataccgc gccacatagc agaactttta aagtgtcat cattggaaaa cgttcttcgg 2940
ggcgaaaaat ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg 3000
caccacaact atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaaacag 3060
gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatac 3120
tcttcctttt tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggataca 3180
tatttgaatg tatttagaaa aataaaca aaaggggttcc gcgcacattt ccccgaaaag 3240
tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa aataggcgta 3300
tcacgaggcc ctttcgtctc gcgcgtttcg gtgatgacgg tgaaaacctc tgacacatgc 3360
agctcccgga gacgggtcac gcttgctctg aagcggatgc cgggagcaga caagcccgctc 3420
agggcgcgct agcgggtgtt ggcggtgtgc ggggctggct taactatgct gcacagatgc 3480
agattgtact gagagtgcac catatgcggt gtgaaatacc gcacagatgc gtaaggagaa 3540
aataccgcat caggcgccat tcgccattca cggctgcgaa ctgttgggaa gggcgatcgg 3600
tgccggcctc ttccgtatta cgccagctgg cgaagggggg atgtgctgca aggcgattaa 3660
gttgggtaac gccagggttt tcccagtcac gacgttgtaa aacgacggcc agtgaattcg 3720
agctgtggaa tgtgtgtcag ttagggtgtg agtgactgca aggtcctccc gaggcgagaa 3780
gtatgcaaaag catgcatctc aattagtcag caaccagggt tggaaagtcc ccaggctccc 3840
cagcaggcag aagtatgcaa agcatgcata ccaattagtc agcaaccata gtcccggccc 3900
taactccgcc catcccgccc ctaactccgc ccagttccgc ccatctctcg 3960
gactaatttt ttttatttat gcagaggccg aggccgcctc ggctctctgag ctattccaga 4020
agtagtgagg aggtcttttt ggaggtctcg taccctcttg cgctaattgct ctgttacagg 4080
tcactaatac catctaagta gttgattcac agtgactgca tatgttgtgt tttacagtat 4140
tatgtagctc gttttttatg caaaatctaa tttaatatat tgatatttat atcattttac 4200
gtttctcggt cagctttttt atactaagtt ggcattataa aaaagcattg cttatcaatt 4260
tggtgcaacg aacaggctac tatcagtc aaataaatca ttatttgatt tcaattttgt 4320
ccactccct gcctctgggg ggcgcg

```

&lt;210&gt; 114

&lt;211&gt; 3166

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; p18attBZeo Plasmid

&lt;400&gt; 114

```

cagttgccgg ccgggtcgcg cagggcgcaac tccgcggccc acggctgctc gccgatctcg 60
gtcatggccg gcccggaggc gtcccggaaag ttctgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccgcg caccacacacc caggccaggg tggtgtccgg caccacctgc 180
tccctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtcgtcc 240
tccacgaagt cccgggagaa cccgagccgg tcggtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgcatgc ctgcaggctc 420
actctagagg atccccgggt accgagctcg aattcgtaat catggtcata gctgtttcct 480
gtgtgaaatt gttatccgct cacaattcca cacaacatac gagccggaag cataaagtgt 540
aaagcctggg gtgcctaatt agtgagctaa ctcacattaa ttgcgttgcg ctactgccc 600
gctttccagt cgggaaacct gtgcgtgccag ctgcattaat gaatcggcca acgcgcgggg 660
agaggcggtt tgcgtattgg gcgctcttcc gcttccctcg cactgactc tcaactgactc 720
gtcgttcggc tgcggcgagg ggtatcagct cactcaaagg cggtaatcag gttatccaca 780
gaatcagggg ataacgcagg aaagaacatg tgagcaaaa gcccagcaaaa ggcaggaac 840
cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gccccctga cgagcatcac 900
aaaaatcgac gctcaagtcg gaggtggcga aaccggacag gactataaag ataccaggcg 960
tttccctctg gaagctccct cgtgcgtctc cctgttccga ccctgcccgt taccggatac 1020
ctgtccgctt ttctcccttc gggaagcggt gcgctttctc atagctcacg ctgtagggtat 1080
ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc cccggttcag 1140
cccgcacgct cgcgcttata cggttaactat cgtcttagt tgaacacgag aagacacgac 1200
ttatcgccac tggcagcagc cactggtaac aggtattgca gagcgaggta tgtaggcggt 1260
gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttggt 1320
atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc 1380

```

-78-

```

aaacaaacca ccgctggttag cgggtgggtttt tttggttgca agcagcagat tacgcgcaga 1440
aaaaaaggat ctcaagaaga tctcttgatc ttttctacgg ggtctgacgc tcagtgggaaac 1500
gaaaactcac gttaaggggat tttggtcatg agattatcaa aaaggatctt cacttagatc 1560
cttttaaaatt aaaaaatgaag ttttaaatca atctaaagta tatatgagta aacttgggtct 1620
gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct atttcgttca 1680
tccatagttg cctgactccc cgtcgtgtag ataactacga tacgggagggg cttaccatct 1740
ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca 1800
ataaaccagc cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc 1860
atccagctta ttaattgttg ccgggaagct gttgtcagaa gttcggccagt taatagtgtg 1920
cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tgggtatggct 1980
tcattcagct ccggttccca acgatcaagg cgagttacat gatcccccat gttgtgcaaa 2040
aaagcggtta gctccttcgg tctccgatc gttgtcagaa gtaagtgtggc cgagtgtta 2100
tcactcatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaatgatgc 2160
ttttctgtga ctgggtgagta ctcaaccaag tcatctgtag aatagtgtat gcggcgaccg 2220
agttgctctt gcccgggcgtc aataccggcg caccctgacg cacatagcag aactttaaaa 2280
gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg 2340
agatccagtt cgatgtaaac cactcgtgca cccaactgat cttcagcatc ttttactttc 2400
accagctgtt ctgggtgagc aaaaaacagg aggcataaat cgcgaaaaaa gggaataagg 2460
gcgacacgga aatgttgaat actcatactc ttcttttttc aatattattg aagcatttat 2520
caggggttatt gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata 2580
ggggttccgc gctcatttcc cgaataagtg ccacctgacg tagttaacaa aaaaaggccc 2640
gccgaagcgg gctttattac caagcgaagc gccattcgcc attcaggctg cgcaactgtt 2700
gggaagggcg atcgggtcgg gctcttcgca tattacgcca gctggcgaaa ggggggatgtg 2760
ctgcaaggcg attaagtgg gtaacgccag ggttttccca gtcacgacgt tgtaaaacga 2820
cgccagtcac gtaatacgac ccttgactag ccttgactag agggtcgacg gtatacagac 2880
atgataagat acattgatga gtttggacaa accacaacta gaatgcagtg aaaaaaatgc 2940
tttatttgtg aaatttgtga tgctattgct ttatttgtaa ccattataag ctgcaataaa 3000
caagttgggg tgggcgaaga actccagcat gagatccccg cgctggagga tcatccagcc 3060
ggcgtcccg aaacgattc cgaagcccaa cctttcatag aaggcggcgg tggaatcgaa 3120
atctcgtagc acgtgtcagt cctgctctc ggccacgaag tgcacg 3160

```

&lt;210&gt; 115

&lt;211&gt; 7600

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; p18attBZeo3'6XHS4eGFP Plasmid

&lt;400&gt; 115

```

cagttgcccgg cggggtcgcg cagggcggaac tcccgcctcc acggctgctc gccgatctcg 60
gtcatggccgg gccgggaggg gtcccgaag tccgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccggt caccacacac caggtcaggg tggttgtccgg caccacctgg 180
tccctggaccg cgctgatgaa caggggtcac tgcgtcccgga ccacaccggc gaagtgcgtcc 240
tccacgaagt cccgggagaa cccgagccgg tcgggtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatcta gttattaata 420
gtaatcaatt acggggtcat tagttcatag cccatatatg gagttccgcg ttacataact 480
tacggtaaat ggcccgcctg gctgaccgcc caacgacccc cgccatttga cgtcaataat 540
gacgtatggt cccatagtaa cgccaatagg gactttccat tgacgtcaat ggggtggacta 600
tttacggtaa actgcccact tggcagtaca tcaagtgtat catatgccaa gtacgcccc 660
tattgacgtc aatgacggta aatggcccg cctggcattat gccagttaca tgaccttatg 720
ggactttcct acttggcagt acatctacgt attagtcac tctccccccc ctccccaccc ccaattttgt 780
tgagccccac gttctgcttc tctctcccc cgtatggggg gggggggggg ggggcgcgcg 840
atttatttat tttttaatta ttttgtgcag cgatgggggc ggccggggcga gggggggggg 900
ccaggcgggg cggggcgggg cgagggggcg ggcggggcga ggcggagagg tgccgcggca 960
gccaatcaga gcggcgcgct ccgtttatgg ccttttatgg cgaggcggcg gggcgggcgg 1020
ccctataaaa agcgaagcgc cgggcggggg ggagtcgctg cgttgccctt gcccgctgcc 1080
ccgctccgcg ccgcctcgcg ccgcccgcgc cggctctgac tgaccgcgtt actcccacag 1140
gtgagcgggg gggacggccc ttctcctcgg ggctgtaatt agcgcttggg ttaatgacgg 1200
ctcgtttctt ttctgtggct gctgaaagg cctaaagggc tccgggaggg ccctttgtgc 1260
ggggggggagc ggctcggggg gtgctgctgc gtgtgtgtgc gtggggagcg ccgcgtgcgg 1320
ccccgcgctg ccggcggtgc tgagcgctgc ggggcggggg gggggctttg tgcgctccgc 1380
gtgtgcgca gggggggcg gtgcccgcgc ggggggggga gtcggggggg gctcgagggg 1440
gaacaaaggc tgcgtgcggg gtgtgtgcgt ggggggggta tgggcgcggc 1500
ggtcgggctg taaccccccc ctgcaccccc cctcccagat ggcccggtt 1560
cgggtgcggg gctcgtgcgc gggcggtggc cggggctcgc cgtgcggggc ggggggtggc 1620

```

ggcaggtggg	ggtgcccggg	ggggcggggg	cgccctcggg	cggggagggg	tcgggggagg	1680
ggcgccggcg	ccccggagcg	ccggcgggctg	tcgagggcg	gcgagccgca	gccattgcct	1740
tttatggtaa	tcgtgcgaga	gggcccaggg	acttcccttg	tcccaaactct	ggcggagccg	1800
aaatctggga	ggcgccggcg	cacccctct	agcgggcgcg	ggcgaagcgg	tgccggcgccg	1860
gcaggaagga	aatgggcggg	gagggccttc	gtgcgtcgcc	gcgcccgcgt	ccccttctcc	1920
atctccagcc	tcggggctgc	cgagggggga	cggtgcctt	cgggggggac	ggggcagggc	1980
gggggttcgg	ttctggcggtg	tgaccggcg	ctctagagcc	tctgctaacc	atgttcatgc	2040
cttcttcttt	ttcctacagc	tccctgggcaa	cgtgtgggtt	gttgtgctgt	ctcatcattt	2100
tggcaaaagaa	ttcgccacca	tgggtgagcaa	gggcgaggag	ctgttcaccg	gggtggtgcc	2160
catcctgggtc	gagctgggacg	gcgacgtaaa	cgggcacaag	ttcagcggtg	ccggcgaggg	2220
cgagggcgat	gccacctacg	gcaagctgac	cctgaagtcc	atctgcacca	ccggcaagct	2280
gcccgtgccc	tggcccaccc	tcgtgaccac	cctgacctac	ggcgtgcagt	gcttcagccg	2340
ctaccccgac	cacatgaagc	agcacgactt	cttcaagtcc	goccatgccc	aagggtacgt	2400
ccaggagcgc	accatcttct	tcaaggacga	cggaactac	aagaccgcg	ccgaggtgaa	2460
gttcgagggc	gacaccctgg	tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	2520
cggcaacatc	ctggggcaca	agctggagta	caactacaac	agccacaacg	tctatatcat	2580
ggccgacaag	cagaagaacg	gcatcaaggt	gaacttcaag	atccgccaca	acatcgagga	2640
cggcagcggtg	cagctcgccg	accactacca	cgagaacacc	cccatcgggc	acggcccgcg	2700
gctgctgccc	gacaaccact	acctgagcac	ccagtcgcgc	ctgagcaaa	accccaacga	2760
gaagcgcgat	cacatgggtcc	tgtgtggagt	cgtgacgcgc	gcccgggatca	ctctcgccat	2820
ggagatccctg	tacaagtaag	aattcactcc	tcaggtgcag	gctgcctatc	agaaggtggg	2880
ggctggtgtg	gccaatgccc	tggctcaca	ataccactga	gatctttttc	cctctgccaa	2940
aaattatggg	gacatcatga	agcccttga	gcatctgact	tctggcta	aaaggaaatt	3000
tattttcatt	gcaatagtgt	gttggaaatt	tttgtgtctc	tcactcgga	ggacataggg	3060
gagggcaaat	catttaaaac	atcagaatga	gtatttgggt	tagagtttgg	caacatatgc	3120
catatgctgg	ctgccatgaa	caaaggtggc	tataaagagg	tcacagtat	atgaaacagc	3180
cccctgctgt	ccattcctta	ttccatagaa	aagccttgac	ttgaggttag	atttttttta	3240
tattttgttt	tgtgttattt	ttttctttta	catccctaaa	attttcctta	catgttttac	3300
tagccagatt	tttctctctc	tccctgactac	tcccagtcac	agctgtccct	cttctcttat	3360
gaagatccct	cgacctgcag	cccaagcttg	catgcctgca	ggtcgactct	agtggatccc	3420
ccgccccgta	tccccaggt	gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	3480
agcgatcccg	tgccaccctc	cccgtgccc	ggctgtcccc	gcacgctgcc	ggctcgggga	3540
tgccgggggg	gcgcgggacc	ggagcgggac	cccggggcg	tcgctgctgc	cccctagcgg	3600
gggagggagc	taattacatc	cctgggggct	ttgggggggg	gctgtcccc	tgagcgggac	3660
cgccggcccc	tatccccag	gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	3720
aaagcgatcc	cgtgccacct	tccccgtgcc	cggtgtgtcc	ccgcacgctg	ccggctcggg	3780
gatgcggggg	gagcgccgga	ccggagcggg	gccccggggc	gctcgctgct	gccccctagc	3840
ggggggagga	cgtaattaca	tccctggggg	ctttgggggg	gggctgtccc	cgtagcgga	3900
tccggggccc	aggtgtctgc	aggtgtctgc	aggtgtctgc	agcagcgaga	agcgttcaga	3960
ggaaagcgat	cccgtgccac	cttccccgtg	ccccgggctg	ccccgcacgc	tgccgggctcg	4020
gggatgcggg	gggagcgccg	gaccgggagc	gagccccggg	cggtcgcgtg	ctgcccccta	4080
gcggggggag	gacgtaatta	catccctggg	ggctttgggg	gggggctgtc	cccgtagcgg	4140
gatccgcggc	cccgtatccc	ccaggtgtct	gcaggtcaca	agagcagcga	gaagcgttca	4200
gaggaagcgc	atcccgtgcc	accttcccc	tgccccgggt	gtccccgcac	gctgcccggc	4260
cggggatgcg	gggggagcgc	cggagccggg	cggagccccg	ggcggtcgcg	tgctgcccc	4320
tagcgggggg	gggacgta	tacatccct	ggggctttgg	ggggggggctg	tccccgtgag	4380
cggatccgcg	gccccgtatc	ccccaggtgt	ctgcaggctc	aaagagcagc	gagaagcgtt	4440
cagaggaag	cgatcccgtg	ccaccttccc	cgtgccccgg	ctgtccccgc	acgctgccc	4500
ctcggggatg	cggggggagc	gcccggaccg	agcggagccc	cgggcggtcc	gctgctgccc	4560
cctagcgggg	gagggacgta	attacatccc	tgggggcttt	gggggggggg	tgtccccgtg	4620
agcgggatcc	cgcccccgta	tccccaggt	gtctgcaggc	tcaaagagca	gcgagaagcg	4680
ttcagaggaa	agcgatcccc	tgccaccttc	cccgtgcccc	ggctgtcccc	gcacgctgcc	4740
ggctcgggga	tgccgggggg	gcgcgggacc	ggagcgggag	ccccggcggc	tcgctgctgc	4800
cccctagcgg	gggaggggac	taattacatc	cctgggggct	ttgggggggg	gctgtcccc	4860
tgagcggatc	cgccggggctg	caggaattcg	taatcatggg	catagctgtt	tccgtgtgta	4920
aattgtttat	cgctcacaat	tccacacaac	atacagcccg	gaagcataaa	gtgtaaagcc	4980
tgggggtgct	aatgagtgag	ctaactcaca	ttaattgcgt	tgcgctcact	gcccgttttc	5040
cagtcgggaa	ccctgtcggtg	ccagctgcac	taatgaatcg	gccaacgcgc	ggggagaggg	5100
ggtttgcgta	ttggggcgctc	ttccgcttcc	tcgctcactg	actcgctgcg	ctcggtcggt	5160
cggtgcggc	gagcgggtatc	agctcactca	aaggcggtaa	tacgggtatc	cacagaatca	5220
gggggataac	caggaaagaa	catgtgagca	aaaggccagc	aaaaggccag	gaaccgtaaa	5280
aaggccgcgt	tgctggcggtt	tttccatagg	ctccgcccc	ctgacgagca	tcacaaaaat	5340
cgacgctcaa	gtcagaggtg	gcgaaacccc	acaggactat	aaagatacca	ggcgtttccc	5400
cctgggaagc	ccctcggtgc	ctctcctgtt	ccgacctgc	cgcttaccgg	atacctgtcc	5460
gcctttctcc	cttcgggaag	cgtggcgctt	tctcatagct	cacgctgtag	gtatctcagt	5520
tcggtgtagg	tcgttcgctc	caagctgggc	tgtgtgcacg	aaccccccg	tcagcccgac	5580
cgctgcgcct	tatccggtaa	ctatcgtctt	gagtcacaac	cggttaagaca	cgacttatcg	5640

-80-

```

ccactggcag cagccactgg taacaggatt agcagagcga ggtatgtagg cgggtgctaca 5700
gagttcttga agtgggtggcc taactacggc tacactagaa ggacagtatt tgggtatctgc 5760
gctctgctga agccagttac cttcggaaaa agagtgggta gctcttgatc cggcacaaca 5820
accaccgctg gtagcgggtgg tttttttggt tgcaagcagc agattacgcg cagaaaaaaa 5880
ggatctcaag aagatccctt gatcttttct acgggggtctg acgctcagtg gaacgaaaac 5940
tcacgttaag ggattttggg catgagatta tcaaaaagga tcttcaccta gatcctttta 6000
aattaaaaat gaagttttta atcaatctaa agtatatatg agtaaacttg gtctgacagt 6060
taccaatgct taatcagtga ggcacctatc tcagcgatct gtctattttc ttcattcata 6120
gttgccctgac tccccgctcg tagataaact acgatacggg aggggttacc atctggcccc 6180
agtgtgcaa tgataccgcg agaccacgcg tcaccggctc cagatttatc agcaataaac 6240
cagccagccg gaagggccga ggcgagaagt ggtcctgcaa ctttatccgc ctccatccag 6300
tctattaatt gttgccggga agctagagta agtagttcgc cagttaatag tttggcgcaac 6360
gttggtgcca ttgctacagg catcgtggtg tcacgctcgt cgtttgggtat ggcttcattc 6420
agctccgggt cccaacgatc aaggcgagtt acatgatccc ccatggttggt caaaaaagcg 6480
gttagctcct cggctcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcactc 6540
atgggttatg cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct 6600
gtgactgggt agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc 6660
tcttgcccg cgtaataacg ggataatacc gcgcacata gcagaacttt aaaagtgtc 6720
atcattggaa aacgtttctt cgtctcaagg ctctcaagg catcttttac tttcaccagc 6780
agttcgatgt aacccactcg tgcacccaac tgatcttcag catcttttac tttcaccagc 6840
gtttctgggt gagcaaaaaa aggaaggcaa aatgccgcaa aaaaggggat aagggcgaca 6900
cggaaatgtt gaatactcat actcttctct tttcaatatt attgaagcat ttatcagggg 6960
tattgtctca tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt 7020
ccgcgcacat ttccccgaaa agtgccacct gacgtagtta acaaaaaaaa gcccgcgaa 7080
gcgggcttta ttaccaagcg aagcgccatt cgccattcag gctgcgcaac tgttgggaag 7140
ggcgatccgg gcgggctctc tcgctattac gccagctggc gaaaggggga tgtgctgcaa 7200
ggcgattaag ttgggtaacg ccagggtttt cccagtcacg acgttgtaaa acgacggcca 7260
gtccgtaata cagctcactt aaggccttga ctagagggtc gacgggtatac agacatgata 7320
agatacattg atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt 7380
tgtgaaatgt gtgatgctat tgcttttatt gtaaccatta taagctgcaa taaacaagtt 7440
gggggtgggt aagaactcca gcatgagatc cccgcgctgg aggatcatcc agccggcgctc 7500
ccggaacacg attccgaagc ccaacctttc atagaaggcg gcgggtggaat cgaaatctcg 7560
tagcacgtgt cagtcctgct cctcggccac gaagtgcacg 7600

```

&lt;210&gt; 116

&lt;211&gt; 7631

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; p18attBZeo5'6XHS4eGFP Plasmid

&lt;400&gt; 116

```

cagttgcccg ccgggtcgcg cagggcgaa ccccccccc acggctgctc gccgatctcg 60
gtcatggccg gcccggaggc gtcccgaag ttctgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccgcg caccacacac caggccaggg tgttgtccgg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtcgtcc 240
tcacgaagt cccgggagaa cccgagccgg tcggtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcacccg aacggcactg gtcaacttgg ccattggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatatac gaattcctgc 420
agcccccgcg atccgctcac ggggacagcc ccccccaaaa gccccagggg atgtaattac 480
gtccctcccc cgctaggggg cagcagcgag ccgcccgggg ctccgctccg gtccggcgct 540
ccccccgcat ccccgagccg gcagcgtgcg gggacagccc gggcacgggg aagggtggcac 600
gggatccgct tcccttgaac gcttctcgct gctctttgag cctgcagaca cctgggggat 660
acggggcccg ggatccgctc acggggacag cccccccca aagccccag ggatgtaatt 720
acgtccctcc cccgctaggg ggcagcagcg agccgcccgg ggctccgctc cggctccggc 780
ctccccccgc atccccgagc cggcagcgtg cggggacagc ccgggacagg ggaagggtgg 840
acgggatcgc tttctctga acgcttctcg ctgctctttg agcctgcaga cacttggggg 900
atacgggggc gcggatccgc tcacggggac agcccccccc caaagcccc agggatgtaa 960
ttacgtccct cccccgtag ggggcagcag cgagccgccc ggggctccgc tccggtccgg 1020
cgctcccccc gcatccccga gccggggaca tgcggggaca ggggaaggta ggggaaggta 1080
gcacgggata gctttcctct gaacgcttct cgctgctctt tgagcctgca gacacctggg 1140
ggatacgggg ccgggatatc gctcacgggg acagccccc cccaaagccc ccagggatgt 1200
aattacgtcc ctccccgct agggggcgag ccggggctcc gctccggtcc gctccggtcc 1260
ggcgctcccc ccgcaccccc gagccggcag cgtgcgggga cagcccgggc acggggaagg 1320
tggcacggga tcgctttcct ctgaacgctt ctcgctgctc tttgagcctg cagacacctg 1380
ggggatacgg ggccgcggat ccgctcacgg ggacagcccc ccccaaaagc ccccgaggat 1440

```



gtaattacgt	ccctcccccg	ctaggggggca	gcagcgagcc	gccccggggct	ccgctccgggt	1500
ccggcgctcc	ccccgcaccc	ccgagccggc	agcgtgcggg	gacagcccg	gcacgggggaa	1560
ggtagggcac	gatcgctttc	ctctgaaccg	ttctcgctgc	tctttgagcc	tgcagacacc	1620
tgggggatac	ggggcgcgcg	atccgctcac	ggggacagcc	ccccccaaa	gccccaggg	1680
atgtaattac	gtccctcccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	1740
gtccggcgct	cccccgcat	ccccgagccg	gcagcgctgc	gggacagccc	gggcacgggg	1800
aaggtggcac	gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttattaata	gtaatcaatt	acggggctcat	1920
tagttcatag	cccatatag	gagttccgcg	ttacataact	tacggtaaat	ggcccgcctg	1980
gctgaccgcc	caacgacccc	cgcccattga	cgtaataat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgcccact	2100
tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagta	tgacctatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tgggtcgagg	tgagccccac	gttctgttc	2280
actctcccca	ctccccccc	ctccccccc	ccaatttgt	atttatttat	tttttaatta	2340
ttttgtgcag	cgatgggggc	gggggggggg	ggggcgcgcg	ccaggcgggg	cggggcgggg	2400
cgagggggcg	ggcgggggcg	ggcgagagg	tgcgggcgca	gccaatcaga	gcggcgcgct	2460
ccgaaggttt	ccctttatgg	cgaggcgcg	gcggcgcgcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtcgctg	cggtgccttc	gccccgtgcc	ccgctccgcg	ccgcctcgcg	2580
ccgccccccc	cggtctgac	tgaccgcgtt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctcctccg	ggctgttaatt	agcgtttggt	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
gcgtgaaagc	cttaaaaggc	tcggggaggg	ccctttgtgc	gggggggagc	ggctcggggg	2760
gtgcgtgcgt	gtgtgtgtgc	gtggggagcg	ccgcgtgcgg	cccgcgctgc	ccggcggtcg	2820
tgagcgctgc	gggcgcggcg	cggggctttg	tgcgctccgc	gtgtgcgcga	ggggagcgcg	2880
gccccggggcg	gtgccccgcg	gtgccccggg	gctgcgaggg	gaacaaaggc	tgctgcgggg	2940
gtgtgtgcgt	gggggggtga	gcaggggggt	tggggcgggc	gggtcggtcg	taaccccccc	3000
ctgacacccc	ctccccgagt	tgctgagcac	ggcccggctt	cggtgcgggg	gctccgtgcg	3060
gggctggcg	cggggctcgc	cggtgcgggc	gggggggtgg	ggcaggtggg	gggtgcgggc	3120
ggggcggggg	cgccctcggg	cggggagggc	tcgggggagg	ggcgcgggcg	ccccggagcg	3180
ccggcggtcg	tcgaggcgcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
gggcgagggg	acttcccttg	tcccaaatct	ggcgagcccg	aaatctggga	ggcgccggcg	3300
cacccccctc	agcgggcgcg	ggcgaagcgg	tgccggcgcc	gcaggaagga	aatggggcg	3360
gagggccttc	gtgcctcgcc	gcggcgccgt	ccccttctcc	atctccagcc	tcggggcgcg	3420
cgacggggga	cggtgcctt	cgggggggac	ggggcgaggg	ggggttcggc	ttctggcggt	3480
tgacggggcg	ctctagagcc	tctgtcaacc	atgttcatgc	cttcttcttt	ttcctacagc	3540
tcttggggcaa	cggtgctggt	gttgtgctgt	ctcatattt	tggcaaagaa	ttcgccacca	3600
tggtgagcaa	gggtagggag	ctgttcaccg	gggtggtgcc	catcctgggt	gagctggacg	3660
gcgacgtaaa	cggccacaag	ttcagcgtgt	ccggcgaggg	cgagggcgat	gccacctacg	3720
gcaagctgac	cctgaagttc	atctgcacca	ccggcaagct	gcccgtgccc	tggcccaccg	3780
tcgtgaccac	ccctgacctc	ggcgtgcagt	gcttcagccg	ctaccccgac	cacatgaagc	3840
agcacgactt	cttcaagttc	gccatgcccc	aaggctacgt	ccaggagcgc	accatcttct	3900
tcaaggacga	cggcaactac	aagaccgcgc	ccaggtgaa	gttcgagggc	gacaccttgg	3960
tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	cggaacatc	ctggggcaca	4020
agctggagta	caactacaac	agccacaacg	tctatatcat	ggccgacaag	cagaagaacg	4080
gcataaaggt	gaacttcaag	atccgccaca	acatcgagga	cggcagcggt	cagctcgccg	4140
accactacca	gcagaacacc	cccacggcg	acggccccgt	gctgctgccc	gacaaccact	4200
acctgagcac	ccagtccgcc	ctgagcaaa	accccaacga	gaagcgcgat	cacatggtcc	4260
tgctggagtt	cgtgaccgcc	gccgggatca	ctctcggcac	ggacgagctg	tacaagtaag	4320
aattcactcc	tcaggtgcag	gctgcctatc	agaagggtgt	ggctgggtgt	gccaatgcc	4380
tggctcacia	ataccactga	gatctttttc	cctctgccaa	aaattatggg	gacatcatga	4440
agccccctga	gcactgtgac	tctggctaat	aaaggaaatt	tattttcatt	gcaatagtgt	4500
gttggaatttt	tttgtgtctc	tcaactcgga	ggacatatgg	gagggcaaat	cattttaaacc	4560
atcagaatga	gtatttggtt	tagagtttgg	caacatatgc	catatgctgg	ctgccatgaa	4620
caaaggtggc	tataaagagg	tcacagttat	atgaaacagc	cccctgctgt	ccatttctta	4680
ttccatagaa	aagccttgac	ttgaggttag	atttttttta	tattttgttt	tgtgttattt	4740
ttttcttttaa	catccctaaa	attttcctta	catgtttttac	tagccagatt	tttccctctc	4800
tcctgactac	tcccagttat	agctgtccct	cttctcttat	gaagatccct	cgacctgcag	4860
cccaagcttg	cctgcctgca	ggctcgactc	cggttaccga	cggttaccga	gctcgaattc	4920
gtaatcatgg	tcatagctgt	ttcctgtgtg	aaattgttat	ccgctcacia	ttccacacaa	4980
catacagacc	ggaagcataa	agtgtaaagc	ctgggggtgc	taatgagtga	gctaactcac	5040
attaatttgc	ttgcgctcac	tgcccgcttt	ccagtcggga	aacctgtcgt	gccacgtcga	5100
ttaatgaatc	ggccaacgcg	cggggagagg	cggtttgcgt	attgggcgct	cttccgcttc	5160
ctcgctcact	gactcgctgc	gctcggtcgt	tcggctgcgg	cgagcggtat	cagctcactc	5220
aaaggcggt	atacggttat	ccacagaatc	aggggataac	gcaggaaaga	acatgtgagc	5280
aaaaggccag	caaaaggcca	ggaaccgtaa	aaaggccgcg	ttgctggcgt	ttttccatag	5340
gctccgcccc	cctgacgagc	atcacaaaaa	tcgacgtcca	agtacagagt	ggcgaaaccc	5400
gacaggacta	taaagatacc	aggcgtttcc	ccctggaagc	tccctcgctg	gctctcctgt	5460

-82-

tccgaccctg	ccgcttaccg	gataacctgc	cgccctttctc	ccttcggggaa	gcgtggcgct	5520
ttctcatagc	tcacgctgta	ggtatctcag	ttcgggtgtag	gtcgttcgct	ccaagctggg	5580
ctgtgtgcac	gaaccccccg	ttcagcccg	ccgctgcgcc	ttatccggt	actatcgtct	5640
tgagtccaac	ccggttaagac	acgacttatc	gccactggca	gcagccactg	gtaacaggat	5700
tagcagagcg	aggatgttag	gcggtgctac	agagtctctg	aagtgggtggc	ctaactacgg	5760
ctacactaga	aggacagtat	ttgggtatctg	cgctctgctg	aagccagtta	ccttcggaaa	5820
aagagttggg	agctcttgat	ccggcaaaaca	aaccaccgct	ggtagcggtg	gtttttttgt	5880
ttgcaagcag	cagattacgc	gcagaaaaaaa	aggatctcaa	gaagatcctt	tgatcttttc	5940
tacgggggtct	gacgctcagt	ggaacgaaaa	ctcacgttaa	gggatttttg	tcatgagatt	6000
atcaaaaagg	atcttcacct	agatcctttt	aaattaaaaa	tgaagtttta	aatcaatcta	6060
aagtatatat	gagtaaacct	ggtctgacag	ttaccaatgc	ttaatcagtg	aggcacctat	6120
ctcagcgatc	tgtctatttc	gttcatccat	agttgcctga	ctccccgtcg	tgtagataac	6180
tacgatacgg	gagggcttac	catctggccc	cagtgcctga	atgataccgc	gagacccacg	6240
ctcaccgggt	ccagatttat	cagcaataaa	ccagccagcc	ggaaggggcg	agcgacaga	6300
tggtcctgca	actttatccg	cctccatcca	gtctattaat	tggtgccggg	aagctagagt	6360
aagtagttcg	ccagtttaata	gtttgcgcaa	cgttgttgcc	attgctacag	gcatcgtggg	6420
gtcacgctcg	tcggttggtg	tggttcatt	cagctccggg	tcccaacgat	caaggcgagt	6480
tacatgattc	cccattgttg	gcaaaaaaag	ggtagctcc	ttcgggtcctc	cgatcgttgt	6540
cagaagtaag	ttggccgcag	tgttatcact	catggttatg	gcagcactgc	ataattctct	6600
tactgtcatg	ccatccgtaa	gatgcttttc	tgtgactggg	gagtactcaa	ccaagtcat	6660
ctgagaatag	gaccgagttg	gaccgagttg	ctcttgccc	gcgtcaatac	gggataatac	6720
cgcgccacat	agcagaactt	taaaagtgtc	catcattgga	aaacgttctt	cggggcgaaa	6780
actctcaagg	atcttacccg	tggtgagatc	cagttcgatg	taaccactc	gtgcacccaa	6840
ctgatcttca	gcatctttta	ctttcaccag	cgtttctggg	tgagcaaaaa	caggaaggca	6900
aaatgccgca	aaaaagggaa	taagggcgac	acggaaatgt	tgaatactca	tactcttcct	6960
ttttcaatat	tattgaagca	tttatcaggg	ttattgtctc	atgagcggat	acatatattg	7020
atgtatttag	aaaaataaac	aaataggggt	tccgcgcaca	tttccccgaa	aagtgccacc	7080
tgacgtagtt	aacaaaaaaa	agcccgcgca	agcgggcttt	attaccaagc	gaagcgccat	7140
tcgccattca	ggctgcgcaa	ctggtgggaa	gggcgatcgg	tcggggcctc	ttcgctatta	7200
cgccagctgg	cgaaaggggg	atgtgctgca	aggcgattaa	gttgggtaac	ggcagggttt	7260
tcccagtcac	gacgttgtaa	aacgacggcc	agtccgtaat	acgactcact	taaggccttg	7320
actagagggg	cgacggtata	cagacatgat	aagatacatt	gatgagtttg	gacaaaaccac	7380
aactagaatg	cagtgaaaaa	aatgctttat	ttgtgaaatt	tgtgatgcta	ttgctttatt	7440
tgtaaccatt	ataagctgca	ataaacaagt	tgggggtggg	gaagaactcc	agcatgagat	7500
ccccgcgctg	gaggatcatc	cagccggcgt	cccggaatac	gattccgaag	cccaaccctt	7560
catagaaggc	ggcgggtggaa	tcgaaatctc	gtagcacgtg	tcagtcctgc	tcctcggcca	7620
cgaagtgcac	g					7631

&lt;210&gt; 117

&lt;211&gt; 4615

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; p18attBZeo6XHS4 Plasmid

&lt;400&gt; 117

cagttgcccg	ccgggtcgcg	cagggcgaa	tcccccccc	acggctgctc	gccgatctcg	60
gtcatggccg	gcccggaggc	gtcccggaa	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggcccg	caccacacc	caggccagg	tggtgtccgg	caccacctgg	180
tcctggaccg	cgctgatgaa	cagggtcacg	tcgtcccga	ccacaccggc	gaagtctgtc	240
tccacgaagt	cccgggagaa	cccagccgg	tcggtccaga	actcgaccgc	tccggcgacg	300
tcgcgcgcgg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgcatgc	ctgcaggctc	420
actctagtgg	atcccccgcc	ccgtatcccc	caggtgtctg	caggctcaaa	gagcagcgag	480
aagcgttcag	aggaaagcga	tcccgtgcc	ccttccccgt	gccccggctg	tccccgcacg	540
ctgccggctc	ggggatgcgg	ggggagcgcc	ggaccggagc	ggagccccgg	gcggctcgct	600
gctgccccct	agcgggggag	ggacgtaatt	acatccctgg	gggctttggg	ggggggctgt	660
ccccgtgagc	ggatccgcgg	ccccgtatcc	cccaggtgtc	tgacggctca	aagagcagcg	720
agaagcgttc	agaggaaagc	gatccccgtg	caccttcccc	gtgccccggc	tgtccccgca	780
cgctgcccgc	tcggggatgc	ggggggagcg	gcggaccgga	gcggagcccc	gggcggctcg	840
ctgctgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	ggggggggct	900
gtccccgtga	gcggatccgc	ggccccgtat	ccccagggtg	tctgcagggt	caaagagcag	960
cgagaagcgt	tcagagga	gcgatcccg	gccaccttcc	ccgtgccccg	gctgtccccg	1020
cacgctgccg	gctcggggat	gcggggggag	cgccggaccg	gagcggagcc	ccgggcgggt	1080
cgctgctgcc	ccctagcggg	ggagggagct	aattacatcc	ctgggggctt	tggggggggg	1140
ctgtccccgt	gagcggatcc	gcggccccgt	atcccccagg	tgtctgcagg	ctcaaagagc	1200



-83-

```

agcgagaagc gttcagagga aagcgatccc gtgccacctt ccccggtgcc gggctgtccc 1260
cgacgcgtgc cggctcgggg atgcgggggg agcgccggac cggagcggag ccccgggcgg 1320
ctcgctgctg cccctagcgc ggggaggggac gtaattacat ccctggggggc tttggggggg 1380
ggctgtcccc gtgagcggat ccgcggcccc gtatccccca ggtgtctgca ggctcaaaga 1440
gcagcgagaa gcgttcagag gaaagcgatc ccgtgccacc ttccccgtgc cgggctgtgc 1500
cccgacgcgt cggcgctcgg ggatgcgggg ggagcgcggg accggagcgg agccccgggc 1560
ggctcgcgtgc tgccccctag cgggggaggg acgtaattac atccctgggg gctttggggg 1620
ggggctgtcc ccgtgagcgg atccgcggcc ccgtatcccc caggtgtctg caggctcaaa 1680
gagcagcgag aagcgttcag aggaaagcga tccccgtcca ccttccccgt gcccgggctg 1740
tccccgcacg ctgcccggctc ggggatgcgg ggggagcgcc ggaccggagc ggagccccgg 1800
gcggctcgct gctgccccct agcgggggag ggaagtaatt acatccctgg gggctttggg 1860
ggggggctgt ccccgtagc ggatccgcgg ggctgcagga attcgtaatc atggctatag 1920
ctgtttcctg tgtgaaattg tttatccgctc tgcctaataa gtgcattaat agccggaagc 1980
ataaagtgtg aagcctgggg gggaaacctg gcgtattggg ggggagcgcc tcacattaat tgcgttgccg 2040
tcactgcccg ctttcagtc gggaaacctg ggcgtatggg taactccagc tgcattaatg aatcgggcaa 2100
cgcgcgggga gaggcggttt gcgtatggg ggggagcgcc cttcctcgct cactgactcg 2160
ctgcgctcgg tcgttcggct taacgcagga ggggagcgcc actcaaaggc ggtaatacgg 2220
ttatccacag ccccgtagc gggaaacctg taacgcagga gagcaaaagg ccagcaaaagg 2280
gccaggaacc gtaaaaaggc cgcgttgctg ggcgttttcc ataggctccg cccccctgac 2340
gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga 2400
taccggcgtg ttccccctgg aagctccctc ggtgcgtctc ctgttccgac cctgcgctt 2460
accggatacc tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca tagctcacgc 2520
tgtaggatc tcagttcggg gtaggtcggt cgccttatcc cgtcccaagc tgggctgtgt gcacgaacct 2580
cccgttcagc cgccttatcc gcttaactatc ttggtcatga gtcttgagtc caaccgggta 2640
agacacgact tatcgccact ggcagcagcc actggttaaca ggattagcag agcgaggtat 2700
gtaggcggtg ctacagagtt cttgaagtgg tggcctaact acggctacac tagaaggaca 2760
gtatttggtg tctgcgctct gctgaagcca gttaccttcg gtaaaaagag tggtagctct 2820
tgatccggca aacaaaccac cgctggtagc ggtggttttt ttgtttgcaa gcagcagatt 2880
acgcgcagaa aaaaaggatc cctttgatct tttctacggg gtctgacgt 2940
cagtggaaacg aaaaactcac gattatcaga gattatcaaa aaggatcttc 3000
acctagatcc ttttaaatta tctaaagtat atatgagtaa atatgagtaa 3060
acttggctcg acagttacca gatctgtcta gatctgtcta 3120
tttcgttcat ccatagttgc ctaactacgat acgggagggc acgggagggc 3180
ttaccatctg gccccagtg cagcgtcacc ggctccagat ggctccagat 3240
ttatcagcaa taaaccagcc tgcaacttta tgcaacttta 3300
tccgcctcca tccagtctat ttcgccagtt ttccgccagtt 3360
aatagtttgc gcaacgttgt ccttcagctc ctccctcggt ctccctcggt 3420
ggatggctt cattcagctc cgttccccaa cctccgatcg ttgtcagaag taagttggcc 3480
ttgtgcaaaa aagcggtag cactcaggtg cttcttactgt catgccatcc 3540
gcagtgttat tttctgtgac cttcttactgt cttcttactgt catgccatcc 3600
gtaagatgct tttctgtgac cttcttactgt cttcttactgt atagtgtatg 3660
cggcgaccga gttgctctg cttcttactgt cttcttactgt acatagcaga 3720
actttaaaag tgctcatcat tttcttactgt cttcttactgt aaggatctta 3780
ccgctgttga gatccagttc tttcttactgt cttcttactgt ttccagcatct 3840
tttacttttc ccagcgtttc tttcttactgt cttcttactgt cgcaaaaatgc 3900
ggaataaggc cgacacggaa tttcttactgt cttcttactgt atattattga 3960
agcattttatc aggggttatg tttcttactgt cttcttactgt ttagaaaaat 4020
aaacaaatag gggttccgcg cttcttactgt cttcttactgt agttaacaaa 4080
aaaaagcccc ccgaagcggg cttcttactgt cttcttactgt tttaggctgc 4140
gcaactgttg ggaagggcga cttcttactgt cttcttactgt ctggcgaaag 4200
gggatgtgc tgcaaggcga ttaagtggg taacgccagg gttttcccag tcacgacgtt 4260
gtaaaacgac ggccagtcg taatacgact cttgactaga gggctgacgg 4320
tatacagaca tgataagata cattgatgag tttggacaaa ccacaactag aatgcagtga 4380
aaaaaatgct ttatttgtga aatttgtgat gctattgctt tatttgtaac cattataagc 4440
tgcaataaac aagttggggg gggcgaagaa ctccagcatg agatccccgc gctggaggat 4500
catccagccg gcgtcccggg aaacgattcc gaagcccaac ctttcataga agggggcggt 4560
ggaatcgaaa tctcgtagca cgtgtcagtc ctgctcctcg gccacgaagt gcacg 4615

```

<210> 118  
 <211> 17384  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pFK161 Plasmid

<400> 118

gcgacacgagg	gagcttccag	ggggaaacgc	ctgggtatctt	tatagtcctg	tcgggggtttc	60
gccacctctg	acttgagcgt	cgatttttgt	gatgctcgtc	agggggggcgg	agcctatgga	120
aaaacgccag	caacgcggcc	tttttacggg	tcttgccctt	ttgctggcct	tttgctcaca	180
tggtctttcc	tgogttatcc	cctgattctg	tggtataaccg	tattaccgcc	tttgagttag	240
ctgataccgc	tcgcccgcgc	cgaacgaccg	agcgcagcga	gtcagtgagc	gaggaagcgg	300
aagagcgtg	acttccgcgt	ttccagactt	tacgaaacac	ggaaaccgaa	gaccattcat	360
gttgttgctc	aggctgcaga	cgttttgcag	cagcagtcgc	ttcacgttcg	ctcgcgtatc	420
gggtgattcat	tctgctaacc	agtaaggcaa	ccccgccagc	ctagccgggt	cctcaacgac	480
aggagcacga	ccgtcgcac	gtggagtaag	cagacatgat	aagatacatt	gatgagtttg	540
gacaaaccac	aactagaatg	cagtgaaaaa	aatgctttat	ttgtgaaatt	tgtgatgcta	600
ttgcttttatt	tgtaaccatt	ataagctgca	ataaacaagt	taacaacaac	aattgcattc	660
attttatgtt	tcaggttcag	ggggagggtg	gggagggttt	ttaaagcaag	taaaacctct	720
acaaatgtgg	tatggctgat	tatgatctct	agtcaaggca	ctatacatca	aatattcctt	780
attaacccct	ttacaaatta	aaaagctaaa	ggtacacaat	ttttgagcat	agttattaat	840
agcagcact	ctatgctgt	gtggagtaag	aaaaaacagt	atgttatgat	tataactggt	900
atgcctactt	ataaaggtta	cagaatattt	ttccataatt	ttcttgata	gcagtgacgc	960
tttttctctt	gtgggtgtaa	tagcaaaagca	agcaagagtt	ctattactaa	acacagcatg	1020
actcaaaaaa	cttagcaatt	ctgaaggaaa	gtccttgggg	tcttctacct	ttcttctctt	1080
ttttggagga	gtagaatgtt	gagagtcagc	agtagcctca	tcactactag	atggcatttc	1140
ttctgagcaa	aacagggttt	cctcattaaa	ggcattccac	cabtgtctcc	attcatcagt	1200
tccatagggt	ggaatctaaa	atacacaac	aattagaatc	agtagtttaa	cacattatac	1260
acttaaaaaa	tttatattta	ccttagagct	ttaaatctct	gtaggtagtt	tgtccaatta	1320
tgtaacacca	cagaagtaag	gttctctcac	aaagatccgg	accaaagcgg	ccatcgtgcc	1380
tcccactcc	tcaggttcgg	gggcatggat	gcgcggatag	ccgctgctgg	tttcttggt	1440
gccgacggat	ttgcaactgcc	ggtagaactc	gcgaggtcgt	ccagcctcag	gcagcagctg	1500
aaccaactcg	cgaggggatc	gagcccgggg	tgggcggaaga	actccagcat	gagatccccg	1560
cgctggagga	tcattccagcc	ggcgtcccg	aaaacgattc	cgaagcccaa	cctttcatag	1620
aaggcggcgg	tggaatcgaa	atctcgtgat	ggcagggttg	gcgtcgcttg	gtcggtcatt	1680
tcgaacccca	gagtcgcgt	cagaagaact	cgtaagaag	gcgatagaag	gcgatgcgt	1740
gcgaatcggg	agcggcgata	ccgtaaagca	cgaggaagcg	gtcagcccat	tcgcccga	1800
gctcttcagc	aatatcacgg	gtagcccaacg	ctatgtcctg	atagcgggtcc	gccacacca	1860
gccggccaca	gtcgatgaat	ccagaaaagc	ggccattttc	caccatgata	ttcggcaagc	1920
aggcatcgcc	atgggtcacg	acgagatcct	cgccgtcggg	atgcgcgcct	tgagcctggc	1980
gaacagttcg	gctggcgaga	gccccgtgat	ctcttcgtcc	agatcatcct	gatcgacaag	2040
accggcttcc	atccgagtac	gtgctcgtcc	gatgcgatgt	ttcgcttggt	ggtcgaatgg	2100
gcaggtagcc	ggatcaagcg	tatgcagccg	ccgcatgca	tcagccatga	tggatacttt	2160
ctcggcagga	gcaaggtag	atgacaggag	atcctgcccc	ggcacttcgc	ccaatagcag	2220
ccagtcctct	cccgtctcag	tgacaacgct	gagcacagct	gcgcaaggaa	cgcccgtcgt	2280
ggccagccac	gatagccgcg	ctgcctcgtc	ctgcagttca	ttcagggcac	cggacaggtc	2340
gggtcttgaca	aaaagaaccg	ggcgccccctg	cgctgacagc	cggaaacagg	cggcatcaga	2400
gcagccgatt	gtctgttggt	cccagtcata	gccgaatagc	ctctccaccc	aagcggccgg	2460
agaacctgcg	tgcaatccat	cttggttcaat	catgcgaaac	gatcctcatc	ctgtctcttg	2520
atcagatctt	gatccccctg	ccctggcggc	ccttggcggc	aagaaagcca	tccagtttac	2580
tttgaggggc	ttcccaacct	taccagaggg	cgccccagct	ggcaattccg	gttcgcttgc	2640
tgtccataaaa	accgcccagt	ctagctatcg	ccatgtaagc	ccactgcaag	ctacctgctt	2700
tctctttgcg	cttgogtttt	cccttgcca	gatagcccag	tagctgacat	tcacccgggg	2760
tcagcacctg	ttctgcggac	tggctttcta	cgtgttcgcg	ttccttttag	agcccttgcg	2820
ccctgagtg	ttgcggcagc	gtgaaaagctt	tttgcaaaa	cctaggcctc	caaaaaagcc	2880
tctcactac	ttctggaata	gctcagaggg	cgaggcgggc	taaataaaaa	aaattagtc	2940
gccaatgggg	ggagaatggg	cggaaactggg	cggagttagg	ggcgggatgg	gaggagttag	3000
ggggcgggact	atgggtgctg	actaattgag	atgcatgctt	tgcatacttc	tgctgctgg	3060
ggagcctggg	gactttccac	acctgggttg	tgactaattg	agatgcatgc	tttgcatact	3120
tctgctgct	ggggagcctg	gggactttcc	acaccctaac	tgacacacat	tccacagccg	3180
gatctgcagg	acccaacgct	gcccgcagatg	cgccgcgtgc	ggctgctgga	gatggcggac	3240
gcgatggata	tggtctgcca	aggggttggtt	tgccgattca	cagttctccg	caagaattga	3300
ttggctccaa	ttcttgagg	ggtgaatccg	ttagcgaggt	gccgcgggct	tccattcagg	3360
tcgaggtggc	ccggctccat	gcaccgcgac	gcaacgcggg	gaggcagaca	aggtataggg	3420
cggcgccctac	aatccatgcc	aacccgttcc	atgtgctcgc	cgaggcgcat	aaatcgccgt	3480
gacgatcagc	ggtccaatga	tcgaagttag	gctggtaaga	gccgcgagcg	atccttgaag	3540
ctgtccctga	tggtcgtcat	ctacctgcct	ggacagcatg	gcctgcaacg	cggcatcccc	3600
atgcgcggcg	aagcgagaag	aatcataatg	gggaaggcca	tccagcctcg	cgtcgcgaa	3660
gccagcaaga	cgtagcccag	cgctcgggc	cgccatgccg	gcgataatgg	cctgcttctc	3720
gccgaaacgt	ttgggtggcg	gaccagtgac	gaaggcttga	gcgagggcgt	gcaagattcc	3780
gaataccgca	agcgacaggc	cgatcatcgt	cgcgctccag	cgaagcggt	cctcgccgaa	3840
aatgacccag	agcgctgccg	gcacctgtcc	tacgagttgc	atgataaaga	agacagtc	3900
aagtgcggcg	acgatagtca	tgccccgcgc	ccaccggaag	gagctgactg	gggtgaaggc	3960
tctcaagggc	atcggctcgac	gctctccctt	atgcgaactcc	tgcattagga	agcagccag	4020

tagtaggttg	aggcgggtga	gcaccgcccgc	cgcaaggaat	ggtgcatgca	aggagatggc	4080
gccccacagt	cccccgccca	cgggcctggc	accataccca	cgccgaaaca	agcgctcatg	4140
agcccgaaat	ggcgagcccg	atcttccccca	tcgggtgatgt	cgccgatata	ggcgccagca	4200
accgcacctg	tggcgccggg	gatgccgggc	acgatgcgtc	cgccgtagag	gatcttggca	4260
gtcacagcat	gcgcataatcc	atgcttcgac	catgcgtca	caaagtaggt	gaatgcgcaa	4320
tgtagtaccc	acatcgatcat	cgctttccac	tgctctcgcg	aataaagatg	gaaaatcaat	4380
ctcatggtaa	tagtccatga	aaatccttgt	attcataaat	cctccaggta	gctatatgca	4440
aattgaaaca	aaagagatgg	tgatctttct	aagagatgat	ggaatctccc	ttcagtatcc	4500
cgatgggtcaa	tgcgctggat	atgggataga	ttgggaatatg	ctgattttta	tgggacagag	4560
ttgcgaactg	ttcccaacta	aaatcatttt	gcacgatcag	cgcactacga	actttaccca	4620
caaatagtca	ggtaatgaat	cctgatataa	agacagggtg	ataaatcagt	cttctacgcg	4680
catcgacgcg	gcacaccgta	gaaagtcttt	cagttgtgag	cctgggcaaa	ccgttaactt	4740
tcggcggtct	tgctgtgcga	caggctcacg	tctaaaagga	aataaatcat	gggtcataaa	4800
attatcacgt	tgctccggcg	ggcgacggat	gttctgtatg	cgctgttttt	ccgtggcgcg	4860
ttgtgtctgt	tgatctggcc	ttctaaatct	ggcacagccg	aattgcgcga	gcttgggttt	4920
gctgaaacca	gacacacagc	aactgaatac	cagaaagaaa	atcactttac	ctttctgaca	4980
tcagaagggc	agaaatttgc	cgttgaacac	ctggtcaata	cgcgtttttg	tgagcagcaa	5040
tattgcgctt	cgatgacgct	tggtgttgag	attgatacct	ctgctgcaca	aaaggcaatc	5100
gacgagctgg	accagcgcat	tcgtgacacc	gtctccttcg	aacttatctg	caatggagtg	5160
tcattcatca	aggacgcgcg	tatcgcaaat	gggtgctatcc	acgcagcggc	aatcgaaaca	5220
cctcgccggg	tgaccaatat	ctacaacatc	agccttggtg	tccagcggtg	tgagccagcg	5280
cagaacaagg	taaccgtcag	tgccgataag	ttcaaagtta	aacctgggtg	tgataccaac	5340
attgaaacgt	tgatcgaaaa	cgcgctgaaa	aacgctgctg	aatgtgcggc	gctggatgtc	5400
acaaagcaaa	tggcagcaga	caagaaagcg	atggatgaac	tggcttctta	tgctccgacg	5460
gccatcatga	tggaaatgttt	ccccgggtgg	gttatctggc	agcagtgcgc	tcgatagtat	5520
gcaatttgata	attattatca	tttgccgggtc	ctttccggcg	atccgccttg	ttacggggcg	5580
gcgacctcgc	gggttttcgc	tatttatgaa	aattttccgg	tttaaggcgt	ttccgttctt	5640
cttcgtcata	acttaatgtt	tttattttaa	ataccctctg	aaaagaaagg	aaacgacagg	5700
tgctgaaagc	gagctttttg	gcctctgtgc	tttcttttct	ctgtttttgt	ccgtggaaatg	5760
aacaatggaa	gtcaacaaaa	agcagctggc	tgacattttc	gggtgcgagta	tcctgtaccat	5820
tcagaactgg	caggaacagg	gaatgcccgt	tctgcgaggc	gggtggcaagg	gtaatgaggt	5880
gctttatgac	tctgcgcgcg	tcataaaaatg	gtatgccgaa	agggatgctg	aaattgagaa	5940
cgaagaagctg	cgcgggagg	ttgaagaagc	gcggcaggcc	agcgaggcag	atccacagag	6000
cgggtgtggt	cgccatgatc	gcgtagtcca	tagtggctcc	aagttagcgaa	gcgagcagga	6060
ctgggcggcg	gcgaagcggt	cggacagtg	tcggagaacg	gggtgcgcata	gaaattgcac	6120
caacgcata	agcgtagca	gcacgcata	tgctgtcgcg	atgctgtcgg	aatggacgat	6180
atcccgcaag	aggcccgga	gtaccggcat	aaccaagcct	atgcctacag	catccagggt	6240
gacggtgcgc	aggatgacga	tgagcgcatt	gttagatttc	atacacggtg	cctgactcgc	6300
ttagcaattt	aactgtgata	aactaccgca	ttaaagctta	tcgatgataa	gcggtcaaac	6360
atgagaatcc	gcggccgctc	ttctcgttct	gccagcgggc	cctcgtctct	ccaccccatc	6420
cgtctgcgcg	tggtgtgtgg	aaggcagggg	tgccggtctc	cgcccccagc	ctgccccgcg	6480
cgcacttttc	tcagtgggtc	gcgtggctcc	tggtgatgtg	tgaggcgccc	gggtgtgccc	6540
tcacgtgttt	cactttggte	gtgtctcgct	tgaccatgtt	cccagagtcg	gtggatgtgg	6600
ccggtggcgt	tgcataccct	tcocgtctgg	tggtgtgcacg	cgctgtttct	tgtaagcgtc	6660
gaggtgctcc	tggagcgttc	cagggtttgtc	tcctaggtgc	ctgcttctga	gctgggtggg	6720
gcgtccccc	ttccctgggt	tgccctcggg	gctccgtctg	gctgtgtgcc	ttcccgtttg	6780
tgtctgagaa	gcccgtgaga	gggggggtcg	ggagagaagg	aggggcaaga	cccccttct	6840
tcgtcgggtg	aggcgccac	cccgcgacta	gtacgcctgt	gcgtagggtc	gggtgctgagc	6900
ggtcgcggct	gggggtggaa	agtttctcga	gagactcatt	gctttcccg	ggggagcttt	6960
gagaggcctg	gctttcgggg	gggacgggtt	gcagggtctc	ccctgtccgc	ggatgctcag	7020
aatgcccttg	gaagagaacc	ttcctgttgc	cgcagacccc	cccgcgcggg	cgcccgcggtg	7080
ttggtcttct	gggttccctg	tggtctcgte	gcactgcate	tctctcggtg	gcccggggctc	7140
gtcgggggtt	tgggtccgtc	ccgcctcag	tgagaaagtt	tccttctcta	gctatcttcc	7200
ggaaaggggtg	cgggtcttct	acggtctcga	gggtctctc	ccgaatggtc	ccctggagggtg	7260
ctcgcctctc	gaccgcctcc	cgcgcgcgca	gcgtttgctc	tctcgtctac	cgcggcccg	7320
ggcctccccg	ctccgagttc	ggggagggat	cacgcggggc	agagcctgtc	tgctcgtcctg	7380
ccgttgtctg	ggagcatgtg	gctcggcttg	tggtgtgggt	ggctggggag	agggtccggt	7440
gcacaccccc	gcgtgcgcgt	actttccctc	cctcctgagg	gcgcgcgtgc	ggagcgggtg	7500
tgggtaggcg	acggtgggct	cccggtccc	caccgctctt	cccgctgcctc	acccgtgcct	7560
tcgctgcgct	cgctccctct	cgtcgcgctc	cacgactttg	gccgctcccc	cgaacggcgcg	7620
ctgcgcgcgt	cgtgtgtgct	tggtgtgtgt	tctcggtgtg	tggtgtgtgt	tcgctcgcgc	7680
cccccttccc	cgcggcagcg	ttcccacggc	tggcgaaatc	gcgggagtc	tccttcccc	7740
cctcgggggtc	gagagggtcc	gtgtctggcg	tcgattgatc	tcgctctcgg	ggacggggacc	7800
gttctgtggg	agaacggctg	ttggccgcgt	cgggcgcgac	gtcggacgtg	gggaccaact	7860
gcccgtcggg	ggtcttcgct	ggtaggcac	gggtgtgtcg	catcggtctc	tctctcgtgt	7920
cgggtgtcgc	tcctcgggct	cccggggggc	cgtcgtgttt	cgggtcgggt	cggcgctgca	7980
gggtgtgggtg	gactgctcag	gggagtgggtg	cagtggtgatt	cccgcgggtt	ttgcctcgcg	8040

tgccctgacc	gggtccgacgc	ccgagcgggtc	tctcgggtccc	ttgtgaggac	ccccctccgg	8100
gaggggcccc	tttcgggccc	ccttgccgctc	gtcgcgggcc	ctcgttctgc	tgtgtcgttc	8160
ccccctcccc	gctcgcgcga	gccgggtcttt	tttctctct	ccccctct	cctctgactg	8220
acccgtggcc	gtgctgtcgg	accccccgca	tgggggcggc	cgggcacgta	cgcgtccggg	8280
cgggtcaccgg	gggtcttgggg	gggggcccag	gggtaagaaa	gtcggctcgg	cgggaggag	8340
gagctgtggt	ttggaggcg	tcccgcccc	gcggccgtgg	cgggtgtctg	cgcggtcttg	8400
gagagggtg	cgtgcgagg	gaaaagggtg	ccccgcgagg	gcaaaggga	agaggctagc	8460
agtgggtcatt	gtcccgcagg	tgtgggtggc	tgttggccga	gggtgcgtctg	gggggctcgt	8520
cgggccccgt	cgtccgtcgg	gaaggcgcgt	gttggggcct	gccggagtgc	cgagggtggg	8580
accctggcgg	tgggattaac	cccgccgcgc	tgtcccgggtg	tggcgggtggg	ggctccgggtc	8640
gatgtctacc	tccctctccc	cgagggtctca	ggccttctcc	gcgcgggctc	tgggcccccc	8700
cctcgttcc	ccctctcgcg	gggttcaagt	cgtcgtcga	cctccccctc	tccgtccttc	8760
catctctcgc	gcaatggcgc	cgcccaggtt	cacgggtgggt	tcgtcctccg	cctccgcttc	8820
tcgcgcgggg	ctggccgctg	tccgggtctct	cctgcccgcac	ccccgttggc	gtgggtcttct	8880
ctcgccggct	ctcgccgactc	ctggctctgc	cgggagggtc	aggggggcttc	cgggtcccc	8940
gacgttgccg	ctcgtctgctg	tgtgcttggg	gggggcccgc	tgcggccctcc	gcccgcctcg	9000
gagccccctg	cgcacccgc	gggtgtcgggt	ttcgcgcgcg	gggtcagttgg	gcccgtggcgt	9060
tgtgtcgcgt	cgggagcgtg	tccgcctcgc	gcgggctaga	cgcgggtgtc	gcccgggtcc	9120
gacgggtggc	ctatccaggg	ctcgcccccg	ccgacccccg	cctgccccgtc	cgggtgggtg	9180
tcgttgggtg	ggggagtga	tggtgtctacc	gggtcattccc	tcccgcggtg	tttgactgtc	9240
tgcgggtgt	cgcgcttctc	tttcgcgcaa	ccccacgc	aaccacccac	cctgtcctcc	9300
cggccccggg	cgggtcgacgt	tccggctctc	ccgatgccga	ggggttcggg	atttgtgccg	9360
gggacggagg	ggagagcggg	taagagaggt	gtcggagagc	tgtcccgggg	cgacgctcgg	9420
tgtggctttg	cgcggtgcgt	gtgctcgcgg	acgggttttg	tcggaccccc	acggggtcgg	9480
tccggccgca	tgcactctcc	cgttccgcgc	gagcgcgcgc	ccggctcacc	cccgggttgt	9540
cctcccgcga	ggctctccgc	cgcgcgcgc	tctcctcct	ctctcgcgct	ctctgtccc	9600
cctggctctg	tcccacccc	gacgtccgc	tcgcgttcc	ttacctgggt	gatcctgcca	9660
ggtagcatat	gcttgtctca	aagattaagc	catgcatgtc	taagtacgca	cggccgggtac	9720
agtgaactg	cgaattggctc	attaaatcag	ttatgggtcc	tttgggtcgt	cgctcctctc	9780
ctaactggat	aactgtggtg	attctagagc	taatacatgc	cgacgggccc	tgacctccct	9840
tcccgggggg	ggatgctgct	atttatcaga	tcaaaaccaa	cccgggtgagc	tccctcccg	9900
ctccggccgg	gggtcggggc	ccggcggtt	gggtgactcta	gataacctcg	ggccgatcgc	9960
acgccccccc	tggcgccgac	gacccattcg	aacgtctgcc	ctatcaactt	tcgatggtag	10020
tcccggtgct	taccattggtg	accacgggtg	acggggaatc	aggggttcgat	tccggagagg	10080
gagcctgaga	aacggctacc	acatccaagg	aaggcagcag	gcgcgcaaat	taccactcc	10140
gcacccgggg	aggttagtgac	gaaaaataac	aatacaggac	tctttcgagg	ccctgtaat	10200
ggaaatgagtc	cactttaaat	cctttaacga	ggatccattg	gagggcaagt	ctgggtgccag	10260
cagccgcgggt	aattccagct	ccaatagcgt	atattaaagt	tgctgcagtt	aaaaagctcg	10320
tagttggatc	ttgggagcgg	gcgggcgggtc	cgccgcgagg	cgagtaccg	cccgtcccc	10380
ccccctgcct	cctcgatgct	cctcgatgct	cttagctgag	tgtcccgcgg	ggcccgaagc	10440
gtttactttg	aaaaaattag	agtgttcaaa	cgaggccgga	gcccgcctgga	taccgcagct	10500
aggaataatg	gaataggacc	gcgggttctat	ttgttgggtt	ttcggaactg	aggccatgat	10560
taagaggggac	ggcggggggc	attcgtattg	cgcgcgtaga	ggtgaaattc	ttggaccggc	10620
gcaagacgga	ccagagcgaa	agcatttgcc	aagaatgttt	tcattaatca	agaacgaaag	10680
tcggagggttc	gaagacgatc	agataccgctc	gtagttccga	ccataaacga	tgccgactgg	10740
cgatgcggcg	gcgttattcc	catgaccgc	cgggcagctt	ccgggaaacc	aaagtctttg	10800
gggtccgggg	ggagtatgggt	tgcaaagctg	aaacttaaa	gaattgacgg	aagggcacca	10860
ccaggagtgg	gcctgcggct	taatttgact	caacacggga	aacctcacc	ggcccgga	10920
cggacaggat	tgacagattg	atagctcttt	ctcgattccg	tgggtgggtg	tgcatggccg	10980
ttcttagttg	gtggagcgat	ttgtctgggt	aattccgata	acgaacgaga	ctctggcatg	11040
ctaactagtt	acgcgacccc	cgagcgggtc	gcgtcccca	acttcttaga	gggacaagtg	11100
gcgttcagcc	acccgagatt	gagcaataac	aggtctgtga	tgcccttaga	tgtccggggc	11160
tgcacgcgcg	ctacactgac	tggctcagcg	tgtgcctacc	ctgcgcgggc	aggcgcgggt	11220
aaccggttga	acccattcgc	tgatggggat	cggggattgc	aattattccc	catgaacgag	11280
gaattcccag	taagtgcggg	tcataagctt	gcgttgatta	agtccctgcc	ctttgtacac	11340
acgcgcgcgt	gctactaccg	attggatgggt	ttagtgaggc	cctcggatcg	gccccgcgg	11400
gggtcggccc	cggccctggc	ggagcgtgta	gaagacgggtc	gaacttgact	atctagagga	11460
agtaaaagtc	gtaacaaggt	ttccgtagggt	gaacctgcgg	aaggatcatt	aaacgggaga	11520
ctgtggaggga	gcggcgcggt	ggccgcgtct	ccccgtcttg	tgtgtgtcct	cgcggggagg	11580
cgcgtgcgtc	cggggctccc	tgcgcgcgt	gtggagcgag	gtgtctggag	tgaggtgaga	11640
gaaggggtgg	gtgggtcgg	tctgggtccg	tctgggacgc	cctccgattt	cccccccc	11700
tccctctctc	ctcgtccggc	tctgacctcg	ccacctacc	gcggcgccgg	ctgctcgcgg	11760
gcgtcttgcc	tctttccgct	cgggtctctc	cgtgtctacg	agggggcggt	cgtcgttacg	11820
gggttttgac	cgttccggg	ggcgttcgggt	cgcgcttggt	cgcgcttggt	tctccggga	11880
cccatccccg	ccgcggctct	ggcttttcta	cgttggctgg	ggcggttgtc	gcgtgtgggg	11940
ggatgtgagt	gtcgcgtgtg	ggctcgcgcg	tcccgatgcc	acgcttttct	ggcctcgcgt	12000
gtcctccccg	ctcctgtccc	gggtacctag	ctgtcgcgtt	ccggcgcgga	ggtttaagga	12060

ccccgggggg	gtcgccctgc	cgccccagg	gtcggggggg	gggtggggccc	gtaggggaagt	12120
cggtcggttcg	ggcggtctctc	cctcagactc	catgaccttc	ctccccccgc	tgccgcggtt	12180
cccgaggcgg	cggtcgtgtg	gggggggtgga	tgtctggagc	ccccctcgggc	gccgtggggg	12240
ccccgaccgc	gcccgcggct	tgccccgattt	ccgcgggtcg	gtcctgtcgg	tgccggtcgt	12300
gggttcccg	gtcgttcccg	tggttttccg	ctcccgaccc	tttttttttc	ctcccccca	12360
cacgtgtctc	gtttcgttcc	tgctggccgg	cctgaggcta	ccccctcggtc	catctgttct	12420
cctctctctc	cggggagagg	agggcggttg	tcgttggggg	actgtgccgt	cgtcagcacc	12480
cgtgagttcg	ctcacacccg	aaataccgat	acgactctta	gcggtggatc	actcggctcg	12540
tcgctcgatg	aagaaccgag	ctagctgcga	gaattaatgt	gaattgcagg	acacattgat	12600
catcgacact	tcgaacgcac	ttgcggcccc	gggttccctcc	cggggctacg	cctgtctgag	12660
cgtcggttga	cgatcaatcg	cgtcacccgc	tgccgtgggt	gctgcgcggc	tgggagtttg	12720
ctcgcagggc	caacccccca	acccgggtcg	ggccctccgt	ctcccgaaagt	tcagacgtgt	12780
gggcgggttgt	cggtgtggcg	cgcgccccc	cgtcgcgag	cctgggtctcc	cccgcgcac	12840
gcgcgtcgcg	gcttcttccc	gctccgcgt	tcccgcctc	gcccgtgcac	cccggctcgt	12900
gcctcgcgtc	ggcgccctcc	ggaccgctgc	ctcaccagtc	tttctcggtc	ccgtgcccc	12960
tgggaaccca	ccgcgcccc	gtggcgcccc	ggggtggcg	cgtccgcac	tgctctggtc	13020
gaggttggcg	gttgaggggtg	tgctgcgc	gaggtgggtg	tcgggtccct	gcggcccgcg	13080
ggttgtcggg	ttggcggtcg	acgagggcg	gtcgttcg	tgcggtgggt	gtctgtgtgt	13140
gtttgggtct	tgcgctgggg	gaggcggggt	cgaccgctcg	cggggttggc	gcggtcgccc	13200
ggcgccgcgc	accctccggc	ttgtgtggag	ggagagcgag	ggcgagaacg	gagagaggtg	13260
gtatcccccg	tgccgttgcg	agggagggtt	tgccgtccc	cgtccgtccg	tccctccctc	13320
cctcggtggg	cgcttccg	ccgcacgcgg	ccgctagggg	cggtcggggc	ccgtggcccc	13380
cgtggctctt	cttcgtctcc	gcttctctct	caccggggcg	gtaccgcgtc	cggcgcggcg	13440
ccgcgggacg	ccgcggcgtc	cgtgcgcgca	ggattccctc	cccgggtgt	tgcgagttcg	13500
gggagggaga	gggcctcgct	gaccggttgc	gtcccggtt	ccctgggggg	gaccggcggt	13560
ctgtgggctg	tgctcccg	gggttgctgt	tgagtaagat	cctccacccc	cgccgcctc	13620
ccctccccgc	gtctcttccc	ggaccccc	cggtctctc	ccggtctcgt	ctcccggtgc	13680
gcccgggtg	gtctcttccc	cgccccctc	ctcgtctctc	tcttcccgcg	gctgggcgcg	13740
tgctccccct	ttctgaccgc	gacctcagat	cagacgtggc	gaccgcgtga	atttaagcat	13800
attagtcagc	ggaggaagaa	aaactaacca	ggattccctc	agtaacggcg	agtgaacagg	13860
gaagagccca	gcgccgaatc	ccgcgcgcgc	gtcgcggcgt	gggaaatgtg	gcgtacggaa	13920
gaccacactc	ccggcgccgc	tcgtgggggg	cccaagtcct	tctgatcgag	gcccagcccc	13980
tggagcggtg	aatgcagccc	gcggcccgga	cgcgccgggc	tcgggtcttc	ccggagtcgg	14040
gttgcttggg	agtcacaacg	aaagcggggtg	gtaaactcca	tctaaggcta	aataccggca	14100
tcagaccgat	gtggaacacg	taccgtaagg	gaaagt tga	aagaactttg	aagagagagt	14160
ctcaagaggg	gtgaaacacg	taagaggtaa	agcgggtggg	tccgcgcagt	ccgcgcggag	14220
gattcaaccc	ggcggcgcgc	gtccggccgt	gcccgggtgt	cccggcggtat	ctttcccgt	14280
ccccgttcc	cccgacccct	ccaccgcgc	gtcgttcccc	tcttccctccc	cgcgtccggc	14340
gctcggcg	gcgggcgcgc	ggggtgggtg	gttgggtggc	cgcgggcggg	gcccgggggtg	14400
gggtcgggcg	gggaccgccc	ccggccggcg	accggccgcc	gcccgggcgca	cttccaccgt	14460
ggcggttgcg	cgcgaccggc	tccgggacgg	ccgggaaggc	ccggtgggga	aggtggctcg	14520
ggggggggcg	cgcgctctcag	ggcgcccgca	accacctcac	cccagtggtt	acagccctcg	14580
ggcgcgctt	tcggggccga	ccggggccga	ggaagccaga	taccgcgtcg	cgcgctctcc	14640
ctctcccccc	gtccgcctcc	cgggcgggcg	tgggggtggg	ggccggggccg	ccccctcccac	14700
ggcgcgaccg	ctctcccacc	ccccctcgtc	gctctctcgt	gggcccgggtg	gggggcgggg	14760
cggactgtcc	ccagtgcgc	ccgggcgtcg	tcgcgcgtc	gggtcccggg	gggaccgtcg	14820
gtcacgcgtc	tcccgacgaa	gcccagcgca	cggggtcggc	ggcgatgtcg	gctaccacc	14880
cgaccgcgtc	tgaaacacgg	accaaggagt	ctaacgcgtg	cgcgagtcag	gggctcgtcc	14940
gaaagccg	gtggcgcaat	gaaggtgaag	ggcccgcgcc	ggggggcccga	ggtgggatcc	15000
cgaggcctct	ccagtccg	gagggcgcac	caccggcccg	tctcgccgc	cgcgccgggg	15060
aggtggagca	cgagcgtacg	cgttaggacc	cgaagatgg	tgaactatgc	ttgggcaggg	15120
cgaagccaga	ggaaactctg	gtggaggtcc	gtagcggtcc	tgacgtgcaa	atcggtcgtc	15180
cgacctgggt	atagggcgga	aagactaatc	gaaccatcta	gtagctgggt	ccctccgaag	15240
tttccctcag	gatagctggc	gctctcgtc	ccgacgtacg	cagttttatc	cggtaaagcg	15300
aatgattaga	ggtcttgggg	ccgaaacgat	ctcaacctat	tctcaaacct	taaatgggta	15360
agaagcccg	ctcgttggcg	tggagccggg	cgtggaatgc	gagtgccctag	tgggccaact	15420
ttggtaagca	gaactggcgc	tgccgggatga	accgaacgcc	gggttaaggc	gcccgatgcc	15480
cagctcatc	agaccgccaga	gttgatag	gttgatag	acagcaggac	gggtggccatg	15540
gaagtcggaa	tccgctaagg	agtgtgtaac	aactcacctg	ccgaatcaac	tagccctgaa	15600
aatggatggc	gctggagcgt	cgggcccata	cccgcccgctc	gcccgcagtcg	gaacggaacg	15660
ggacgggagc	ggcgcggaat	tcttgaagac	gaaagggcct	cgtgatacgc	catattttat	15720
aggttaatgt	catgataata	atgggtttctt	agacgtcagg	tggcactttt	cggggaaatg	15780
tgcgcggaac	ccctatttgt	ttatttttct	aaatacatc	aaatatgtat	ccgctcatga	15840
gacaataacc	cttcaataat	cttcaataat	attgaaaaag	gaagagtatg	agtattcaac	15900
atttccgtgt	cgcccttatt	cccttttttg	cggcattttg	cttccctgtt	ttgctcacc	15960
agaaacgctg	gtgaaagtaa	aagatgctga	agatcagttg	ggtgcacgag	tgggttacat	16020
cgaactggat	ctcaacagcg	gtaagatcct	tgagagttt	cgccccgaag	aacgttttcc	16080

-88-

```

aatgatgagc acttttaaa gttctgctatg tggcgcggtg ttatcccgtg ttgacgcggg 16140
gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggttg agtactcacc 16200
agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca gtgctgccat 16260
aaccatgagt gataaacttg cggccaactt acttctgaca acgatcggag gaccgaagga 16320
gctaaccgct tttttgcaca acatggggga tcatgtaact cgccttgatc gttgggaacc 16380
ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg cagcaatggc 16440
aacaacgttg cgcaaactat taactggcga actacttact ctgctctccc ggcaacaatt 16500
aatagactgg atggaggcgg ataaagtgc aggaccactt ctgcgctcgg cccttcggc 16560
tggctgggtt attgctgata aatctggagc cggtgagcgt ggggtctcgg gtatcatatg 16620
agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacga cggggagtca 16680
ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac tgattaagca 16740
ttggttaactg tcagaccaag tttactcata tatacttttag attgatttaa aacttcattt 16800
ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca aaatccctta 16860
acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag gatcttcttg 16920
agatcccttt tttctgcgct taatctgctg ctgcaaaaca aaaaaaccac cgctacccag 16980
ggtggtttgt ttgccggatc aagagctacc aactcttttt ccgaaggtaa ctggcttcag 17040
cagagcgcag ataccaaata ctgtccttct agtgtagccg tagttaggcc accacttcaa 17100
gaactctgta gcaccgccta catacctcgc tctgctaata ctggtaccag tgggtgctcg 17160
cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac cggataaagg 17220
gcagcggctg ggctgaacgg ggggttcgtg cacacagccc agcttggagc gaacgacct 17280
caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc cgaagggaga 17340
aaggcggaca ggtatccggt aagcggcagg gtcggaacag gaga 17384

```

&lt;210&gt; 119

&lt;211&gt; 2814

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pLITMUS38 Plasmid

&lt;400&gt; 119

```

gttaactacg tcagggtggca cttttcgggg aatgtgctgc ggaaccccta tttgtttatt 60
tttctaaata cattcaaata tgtatccgct catgagacaa taacctgat aaatgcttca 120
ataatattga aaaaggaaga gtatgagtag tcaacatttc cgtgtcgccc ttattccctt 180
ttttgcggca ttttgccttc ctgtttttgc tccccagaa acgctggtga aagtaaaaga 240
tgctgaagat cagttgggtg cacgagtggg ttacatcgaa ctggatctca acagcggtaa 300
gatccttgag agttttcgcc ccgaagaacg ttctccaatg atgagcactt ttaaagtctt 360
gctatgtggc gcggtattat cccgtgttga cgcggggcaa gagcaactcg gtcgcgcgat 420
acactattct cagaatgact tggttgagta ctaccagtc ctacaaaagc atcttacgga 480
tggcatgaca gtaagagaat tatgcagtgc tgccataacc atgagtgata acactgcggc 540
caacttactt ctgacaacga tcggaggacc gaaggagcta accgcttttt tgcacaaact 600
ggggatcat gtacttcgct tggatcgttg ggaaccggag ctgaatgaag ccataccaaa 660
cgacgagcgt gacaccacga tgctgtgagc aatggcaaca acgttgcgca aactatatac 720
tggcgaaact ctactcttag cttcccggca acaattaata gactggatgg aggcggataa 780
agttgcagga ccacttctgc gctcggccct tccggctggc tggtttattg ctgataaatc 840
tggagccggt gagcgtgggt ctgcgggtat cattgcagca ctggggccag atggttaagg 900
ctcccgatc gtatgtatct acacgacggg gagtcaggca actatggatg aacgaaatag 960
acagatcgct gagatagggt cctcactgat taagcattgg taactgtcag accaagttta 1020
ctcatatata ctttagattg atttaccocg gttgataatc agaaaagccc caaaaacagg 1080
aagattgtat aagcaaatat ttaaattgta aacgttaata ttttgtaaa attcgcgtta 1140
aatttttggt aaatcagctc attttttaac caataggccg aaatcggcaa aatcccttat 1200
aaatcaaaag aatagcccgga gatagggttg agtggtgttc cagtttggaa caagagtcca 1260
ctattaaaga acgtggactc caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc 1320
ccactacgtg aaccatcacc caaatcaagt tttttggggc cgaggtgccc taaagcacta 1380
aatcggaacc ctaaaggagag cccccgattt agagcttgac ggggaaagcg aacgtggcga 1440
gaaagggaag gaagaaaagc aaaggagcgc gcgctagggc gctggcaagt gtagcggta 1500
cgctgcgcgt aaccaccaca ccgcgcgcgc ttaactgcgc gctacagggc gctacaaaag 1560
atctaggtga agatcctttt tgataatctc atgacaaaaa tcccttaacg tgagttttcg 1620
ttccactgag cgtcagaccc cgtagaaaaa atcaaaggat cttcttgaga tccctttttt 1680
ctcgcggtaa tctgtgctt gcaaacaaaa taccagcggt ggtttgtttg 1740
ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata 1800
ccaaatactg ttcttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca 1860
ccgctacat acctcgctct gctaactcgt ttaccagtgg ctgctgccag tggcgataag 1920
tcgtgtctta ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc 1980
tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 2040
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aaggagaaaa ggcggaacag 2100

```



-89-

tatccggtaa	gcggcagggg	cggaacagga	gagcgacga	gggagcttcc	agggggaaac	2160
gcctgggtatc	tttatagtc	tgctcgggtt	cgccacctct	gacttgagcg	tcgattttttg	2220
tgatgctcgt	cagggggg	gagcctatgg	aaaaacgcca	gcaacgcggc	ctttttacgg	2280
ttcctggcct	tttgctggcc	ttttgctcac	atgtaatgtg	agttagctca	ctcattaggc	2340
accccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tgagggaattg	tgagcggata	2400
acaatttcac	acaggaaca	gctatgacca	tgattacgcc	aagctacgta	atacgactca	2460
ctagtggggc	ccgtgcaatt	gaagccggct	ggcgccaagc	ttctctgcag	gatatactgga	2520
tccacgaatt	cgctagcttc	ggcgtgacg	cgtctccgga	tgtagaggca	tgcgctgacc	2580
ctctagtcaa	ggccttaagt	gagtcgtatt	acggactggc	cgctggttta	caacgctgtg	2640
actgggaaaa	ccctggcggt	acccaactta	atcgcttgc	agcacatccc	cctttcgcca	2700
gctggcgtaa	tagcgaagag	gcccgcaccg	atcgcccttc	ccaacagttg	cgcgccctga	2760
atggcgcaatg	gcgcttcgct	tggtataaaa	gcccgccttcg	gcgggctttt	tttt	2814

&lt;210&gt; 120

&lt;211&gt; 2847

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pLIT38attB Plasmid

&lt;400&gt; 120

gttaactaag	tcagggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	tttgtttatt	60
tttctaata	cattcaata	tgatccgct	catgagacaa	taaccctgat	aaatgcttca	120
ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtgcgcc	ttattccctt	180
ttttgcgga	ttttgccttc	ctgtttttgc	tcacccagaa	acgctgggtga	aagtaaaaga	240
tgctgaagat	cagttgggtg	cacgagtggtg	ttacatcgaa	ctggatctca	acagcggtaa	300
gatccttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagttct	360
gctatgtggc	gcgggtattat	cccgtgttga	cgccggggcaa	gagcaactcg	gtcgccgcat	420
acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	480
tgccatgaca	gtaagagaat	tatgcagtgc	tgccataacc	atgagtata	acactgcggc	540
caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	tgcaacaat	600
gggggatcat	gtaactcgcc	ttgatcgttg	ggaacgggag	ctgaatgaag	ccataccaaa	660
cgacgagcgt	gacaccacga	tgccgttagc	aatggcaaca	acgttgcgca	aactattaac	720
tgccgaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	aggcggtata	780
agttgcagga	ccacttctgc	gctcgccctt	tcgggtggc	tggtttattg	ctgataaatc	840
tgagcgcggt	gagcggtggg	ctcgcggtat	cattgcagca	ctggggccag	atggtaagcc	900
ctcccgatc	gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	960
acagatcgct	gagataggtg	cctcactgat	taagcattgg	taactgtcag	accaagtcta	1020
ctcatatata	ctttagattg	atttaccccg	gttgataatc	agaaaagccc	caaaaacagg	1080
aagattgtat	aagcaaatat	ttaaattgta	aacgttaata	ttttgttaaa	attcgcggtta	1140
aattttttgt	aaatcagctc	attttttaac	caataggccg	aaatcggcaa	aatcccttat	1200
aaatcaaaag	aatagcccga	gatagggttg	agtgttgttc	cagtttgga	caagagtcca	1260
ctattaaaga	acgtggactc	caacgtcaaa	ggcggaaaaa	ccgtctatca	ggcgcatggc	1320
ccactacgtg	aaccatcacc	caaatcaagt	tttttggggg	cgaggtgccc	taaaagccta	1380
aatcggaacc	ctaaagggag	ccccgatttt	agagcttgac	ggggaaaagc	aacgtggcga	1440
gaaaggaagg	gaagaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcggtca	1500
cgctgcgcgt	aaccaccaca	cccgcgcgcg	ttaatgcgcc	gctacagggc	gcgtaaaagg	1560
atctagggtga	agatcctttt	tgataatctc	atgaccaaaa	tcctttaacg	tgagttttcg	1620
ttccactgag	cgctcagacc	cgtagaaaag	atcaaaggat	cttcttgaga	tccttttttt	1680
ctgcgcgtaa	tctgctgctt	gcaaacaaaa	aaaccaccgc	taccagcggt	ggtttgtttg	1740
ccggatcaag	agctaccaac	tctttttccg	aaggtaactg	gcttcagcag	agcgagata	1800
ccaaataactg	ttcttctagt	gtagccgtag	ttaggccacc	acttcaagaa	ctctgtagca	1860
ccgcctacat	acctcgctct	gctaactctg	ttaccagtgg	ctgctgccag	tgccgataag	1920
tcgtgtctta	ccgggttgga	ctcaagacga	tagttaccgg	ataaggcgca	gcggtcgggc	1980
tgaacggggg	gttcgtgcac	acagcccagc	ttggagcgaa	cgacctacac	cgaactgaga	2040
tacctacagc	gtgagctatg	agaaagcgcc	acgcttcccg	aaggggagaaa	ggcggacagg	2100
tatccggtaa	gcggcagggg	cggaacagga	gagcgacga	gggagcttcc	agggggaaac	2160
gcctgggtatc	tttatagtc	tgctcgggtt	cgccacctct	gacttgagcg	tcgattttttg	2220
tgatgctcgt	cagggggg	gagcctatgg	aaaaacgcca	gcaacgcggc	ctttttacgg	2280
ttcctggcct	tttgctggcc	ttttgctcac	atgtaatgtg	agttagctca	ctcattaggc	2340
accccaggct	ttacacttta	tgcttccggc	tcgtatgttg	tgagggaattg	tgagcggata	2400
acaatttcac	acaggaaca	gctatgacca	tgattacgcc	aagctacgta	atacgactca	2460
ctagtggggc	ccgtgcaatt	gaagccggct	ggcgccaagc	ttctctgcag	gatatactgga	2520
tgcttttttta	tactaacttg	agcgaaatct	ggatccacga	attcgctagc	ttcgcccggtg	2580
acgcgtctcc	ggatgtacag	gcatgcgtcg	accctctagt	caaggcctta	agtgagtcgt	2640
attacggact	ggccgtcgtt	ttacaacgtc	gtgactggga	aaaccctggc	gttacccaac	2700

-90-

ttaatcgccct	tgcagcacat	ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	2760
ccgatcgccc	ttcccaacag	ttgcgcagcc	tgaatggcga	atggcgcttc	gcttggtaat	2820
aaagcccgct	tcggcgggct	ttttttt				2847

&lt;210&gt; 121

&lt;211&gt; 4223

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pLIT38attBBSRpolyA2 Plasmid

&lt;400&gt; 121

accatgaaaa	cattttaacat	ttctcaacaa	gatctagaat	tagtagaagt	agcgacagag	60
aagattacaa	tgctttatga	ggataataaa	catcatgtgg	gagcgggcaat	tcgtacgaaa	120
acaggagaaa	tcatctcggc	agtacatatt	gaagcgtata	taggacgagt	aactgtttgt	180
gcagaagcca	ttgcgattgg	tagtgacgtt	tcgaatggac	aaaaggattt	tgacacgatt	240
gtagctgtta	gacaccctta	ttctgacgaa	gtagatagaa	gtattcggagt	ggtaagtcct	300
tgtggtatgt	gtagggagtt	gatttcagac	tatgcaccag	attgttttgt	gttaatagaa	360
atgaatggca	agtttagtcaa	aactacgatt	gaagaactca	ttccactcaa	atatacccca	420
aattaaaagt	tttaccatac	caagcctggc	tgctgcctga	ggctgggacga	cctcgcggag	480
ttctaccggc	agtgcacatc	cgtcggcatc	caggaaacca	gcagcggcta	tcgcgcgcatc	540
catgcccccg	aactgcagga	gtgggggaggc	acgatggccg	ctttgggtccg	gatcttttgt	600
aaggaaacctt	acttctgtgg	tgtgacataa	ttggacaaac	tacctacaga	gattttaaagc	660
tctaaggtaa	atataaaatt	tttaagtgtg	taatgtgtta	aactactgat	tctaatttgt	720
tgtgtatttt	agattccaac	ctatggaact	gatgaatggg	agcagtgggtg	gaatgccttt	780
aatgagaaa	acctgttttg	ctcagaagaa	atgccatcta	gtgatgatga	ggctactgct	840
gactctcaac	attctactcc	tcacaaaaag	aagagaaagg	tagaagaccc	caaggacttt	900
ccttcagaat	tgctaagttt	tttgagtcac	gctgtgttta	gtaatagaac	tcttgcttgc	960
tttgctattt	acaccacaaa	ggaaaaagct	gcactgctat	acaagaaaaa	tatggaaaaa	1020
tattctgtaa	cctttataag	taggcataac	agttataatc	ataacatact	gttttttctt	1080
actccacaca	ggcctagagt	gtctgctatt	aataactatg	ctcaaaaaat	gtgtaccttt	1140
agctttttta	tttgtaaaag	ggtttaataag	gaatatttga	tgtatagtgc	ccttgactaga	1200
gatcataatc	agccatacca	catttgtaga	ggttttactt	gctttaaaaa	acctcccaca	1260
cctccccctg	aacctgaaac	ataaaaatgaa	tgcaatttgt	gttggttaact	tgttttattgc	1320
agcttataat	ggttataaat	aaagcaatag	catcacaaat	ttcacaaaata	aagatccaga	1380
tttcgctcaa	gttagtataa	aaaagcaggg	ttcaatcctg	cagagaagct	tgccgcgcagc	1440
cggtctcaat	tgacggggcc	ccactagtga	gtcgtattac	gtagcttggc	gtaatcatgg	1500
tcatagtctg	ttcctgtgtg	aaattgttat	ccgctcacaa	ttccacacaa	catacgagcc	1560
ggaagcataa	agtgtaaaag	ctgggggtgcc	taatgagtga	gctaactcac	attacatgtg	1620
agcaaaaagg	cagcaaaaagg	ccaggaaccg	taaaaaggcc	gcgttgctgg	cgttttttcca	1680
taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtccga	ggtaggcgaaa	1740
cccgacagga	ctataaaagt	accaggcggt	tccccctgga	agctccctcg	tgccgctctcc	1800
tggtccgacc	ctgccgctta	ccggataacct	gtccgccttt	ctcccttcgg	gaagcgtggc	1860
gctttctcat	agctcacgct	gtaggtatct	cagttcgggtg	taggtcgttc	gctccaagct	1920
gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	1980
tcttgagtc	aacccggtaa	gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	2040
gattagcaga	gcgaggtatg	taggcgggtg	tacagagttc	ttgaagtgg	ggcctaacta	2100
cggtacact	agaagaacag	tatttggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	2160
aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	gtgggttttt	2220
tggttgcaag	cagcagatta	cgcgcagaaa	aaaaggatct	caagaagatc	ctttgatctt	2280
ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	taagggattt	tggtcatgag	2340
attatcaaaa	aggatcttca	cctagatcct	tttacgcgcc	ctgtagcggc	gcattaaagc	2400
cggcggggtg	ggtgggttac	cgcagcgtga	ccgctacact	tgccagcgcc	ctagcgcgcc	2460
ctcctttcgc	ttctttccct	tcctttctcg	ccaggttcgc	tttccccgtc	aagctctaaa	2520
tcggggggctc	cctttagggt	tccgatttag	tgctttacgg	cacctcgacc	ccaaaaaact	2580
tgatttgggt	tatgggtcac	gtagtgggcc	atcgccctga	tagacgggtt	ttcgcccttt	2640
gacgttggag	gccaggttct	ttaatagtgg	actcttgttc	caaactggaa	caacactcaa	2700
ccctatctcg	ggctattctt	ttgatttata	agggattttg	ccgatttcgg	cctatttggt	2760
aaaaaatgag	ctgatttaac	aaaaatttaa	cgcgaatttt	aacaaaatat	taacgtttac	2820
aatttaataa	tttgcttata	caatcttcc	gtttttgggg	cttttctgat	tatcaaccgg	2880
ggtaaatcaa	tctaaagtat	atatgagtaa	acttggtctg	acagttacca	atgcttaatc	2940
agtgaggcac	ctatctcagc	gatctgtcta	tttcgctcat	ccatagtgc	ctgactcccc	3000
gtcgtgtaga	taactacgat	acgggagggc	ttaccatctg	gccccagtg	tgcaatgata	3060
ccgcgagacc	cacgctcacc	ggctccagat	ttatcagcaa	taaaccagcc	agccgggaag	3120
gccgagcgca	gaagtggctc	tgcaacttta	tccgectcca	tccagtctat	taattgttgc	3180
cgggaagcta	gagtaagtag	ttcgccagtt	aatagtttgc	gcaacgttgt	tgccattgct	3240



-91-

```

acaggcatcg tgggtgtcacg ctcgtcgcttt ggtatggctt cattcagctc cggttcccaa 3300
cgatcaaggc gagttacatg atcccccatg ttgtgcaaaa aagcgggttag ctccctcggt 3360
cctccgatcg ttgtcagaag taagttggcc gcagtgttat cactcatggt tatggcagca 3420
ctgcataatt ctcttactgt catgccatcc gtaagatgct tttctgtgac tgggtgagtac 3480
tcaaccaagt cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgctca 3540
acacgggata ataccgcgcc acatagcaga actttaaaaag tgctcatcat tggagaacgt 3600
tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaaccc 3660
actcgtgcac ccaactgac ttcagcatct tttactttca ccagcgtttc tgggtgagca 3720
aaaacaggaa ggcaaaaatgc cgcaaaaaaag ggaataaggcg cgacacggaa atggttgaata 3780
ctcatactct tccttttttca atattattga agcattttatc aggggttattg tctcatgagc 3840
ggatacatat ttgaatgtat ttagaaaaaat aaacaaatag gggttccgcg cacatttccc 3900
cgaaaagtgc cacctgacgt agttaacaaa aaaaagcccg ccgaagcggg ctttattacc 3960
aagcgaagcg ccattcgcca ttcaggctgc gcaactgttg ggaagggcga tcggtgcggg 4020
cctcttcgct attacgccag ctggcgaaaag ggggatgtgc tgcaaggcga ttaagtggg 4080
taacgccagc gttttcccag tcacgacggt gtaaaacgac ggccagtcgc taatacgact 4140
cacttaaggc cttgactaga gggtcgacgc atgcctgtac atccggagac gcgtcacggc 4200
cgaagctagc gaattcgtgg atc 4223

```

&lt;210&gt; 122

&lt;211&gt; 2686

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pUC18 Plasmid

&lt;400&gt; 122

```

tcgcgcgctt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60
cagcttgtct gtaagcggat gccgggagca gacaaagccc tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc 240
attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcggggc tcttcgctat 300
tacgccagct ggcgaaaggg gatgtgctg caaggcgatt aagttgggta acgccagggt 360
tttcccagtc acgacgttgt aaaaacgacgg ccagtgccaa gcttgcacgc ctgcaggtcg 420
actctagagg atccccgggt accgagctcg aattcgtaat catggtcata gctgtttcct 480
gtgtgaaatt gtatccgct cacaattcca cacaacatac gagccggaag cataaagtgt 540
aaagcctggg gtgcctaata gtgagctaa ctcacattaa ttgcgttgcg ctcactgccc 600
gctttccagt cgggaaacct gtgcgtgccc ctcacattaa gaatcggcca acgcgcgggg 660
agaggcgggt tgcgtattgg gcgctcttcc gcttctcgc tcaactgactc gctgcgctcg 720
gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtataacg gttatccaca 780
gaatcagggg ataacgcagg aaagaacatg tgaacaaaag gccagcaaaa ggccagggaac 840
cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gccccctga cgagcatcac 900
aaaaatcgac gctcaagtca gagggtggcg aaccgcagac gactataaag ataccaggcg 960
tttcccctg gaagctcccc cgtgcgctct cctgttccga ccctgcccgt taccggatac 1020
ctgtccgcct ttctccctc gggaaagcgt gcgctttctc atagctcacg ctgtaggtat 1080
ctcagttcgg tctaggtcgt tcgctccaa cttgggtgtg tgcaacgaac ccccggtcag 1140
cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaacccggg aagacacgac 1200
ttatcgccac tggcagcagc cactggtaac aggttagca gagcgaggta tgtaggcgg 1260
gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttgg 1320
atctgcgctc tgctgaagcc agttaccttc ggaaaaaagag ttggtagctc ttgatccggc 1380
aaacaaacca ccgctggtag cgggtggttt tttgtttgca agcagcagat tacgcgcaga 1440
aaaaaaggat ctcaagaaga tcctttgatc tttctacgg ggtctgacgc tcagtggaaac 1500
gaaaactcac gttaagggat ttttgtcatg agattatcaa aaaggatctt cacttagatc 1560
cttttaaat aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggtct 1620
gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct attcgttca 1680
tccatagttg cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct 1740
ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggtccaga tttatcagca 1800
ataaaccagc cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc 1860
atccagctca ttaattgttg ccgggaagct agagtaagta gttccggcag taatagtttg 1920
cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggtagggc 1980
tcattcagct ccggttccca acgatcaagg cyagtacat gatcccccac gttgtgcaaa 2040
aaagcggtta gctccttcgg tcctccgatc gttgtcagaa gtaagttggc cgcagtggtta 2100
tcactcatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaaagatgc 2160
tttctgtga ctggtgagta ctcaaccaag tcatctgag aatagtgat gggcgagccg 2220
agttgctctt gcccgcgctc aatacgggat aataccgcgc cacatagcag aactttaaaa 2280
gtgctcatca ttggaaaaacg ttcttcgggg cgaaaactct accgctgttg 2340
agatccagtt cgtagttaacc cactcgtgca cccaactgat cttcagcatc ttttactttc 2400

```

-92-

accagcgttt	ctgggtgagc	aaaaacagga	aggcaaaatg	ccgcaaaaaa	gggaataagg	2460
gcgacacgga	aatgttgaat	actcatactc	ttcctttttc	aatattattg	aagcatttat	2520
cagggttatt	gtctcatgag	cggatacata	tttgaatgta	tttagaaaaa	taaacaaata	2580
ggggttccgc	gcacatttcc	ccgaaaagtg	ccacctgacg	tctaagaaac	cattattatc	2640
atgacattaa	cctataaaaa	taggcgtatc	acgaggccct	ttcgtc		2686

&lt;210&gt; 123

&lt;211&gt; 8521

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pCXeGFPattB(6xHS4)2 Plasmid

&lt;400&gt; 123

tacggggcgg	gggatccact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccataat	ggagttccgc	gttacataac	ttacgggtaaa	tggcccgcct	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtac	240
atcaagtgt	tcatatgcc	agtacgcccc	ctattgacgt	caatgacggt	aaatggcccc	300
cctggcatta	tgcccagtac	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tatttagtcat	cgctattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctcccc	420
atctcccccc	cctccccacc	cccaattttg	tatttattta	ttttttaatt	attttgtgca	480
gcgatggggg	cggggggcgc	ggggggcgcg	gccaggcgcg	gcggggcgcg	gcgaggggcg	540
ggggcgggcg	aggcgagag	gtgcgggcgc	agccaatcag	agcgggcgcg	tccgaaagtt	600
tccttttatg	gcgagggcgc	ggcgggcgcg	gcccataaaa	aagcgaagcg	cgcgggcggg	660
gggagtcgct	gcgttgccct	cgccccgtgc	cccgcctcgc	gccgcctcgc	gccgcccgcc	720
ccggctctga	ctgaccgcgt	tactcccaca	gggtgagcg	cgggacggcc	cttctcctcc	780
gggctgtaat	tageccttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgcgtgaaag	840
ccttaaagg	ctccgggagg	gccctttgtg	cgggggggag	cggctcgggg	ggtgcgtgcg	900
tgtgtgtgtg	cgtggggagc	gccgcgtcgc	gcccgcgctg	ccggcgcgct	gtgagcgctg	960
cgggcgggcg	gcggggcctt	gtgcgctcgc	cgtgtgcgcg	aggggagcgc	ggcggtgggg	1020
gggtgccccg	gggtgcgggg	ggctgcgagg	ggaacaaaag	ctgcgtgcgg	ggtgtgtgct	1080
tggggggggtg	agcagggggg	gtggggcgcg	cgggtcgggt	gtaaccccc	ctctgacccc	1140
cctccccgag	ttgctgagca	cgccccggct	tcgggtgcgg	ggctccgtgc	ggggcggtgg	1200
gcggggctcg	cctgcgccgg	cgggggggtg	cggcaggtgg	gggtgccggg	cggggcgggg	1260
cgccctcggg	ccggggaggg	ctcgggggag	ggggcgggcg	gccccggagc	gcccggcgct	1320
gtcgagggcg	ggcgagccgc	agccattgcc	ttttatggta	atcgtgcgag	agggcgaggg	1380
gacttccctt	gtcccaaatc	tggcggagcc	gaaatctggg	aggcgccgcc	gcacccctcc	1440
tgcggggcgc	gggcgaagcg	gtgcggcgcc	ggcaggaagg	aaatggggcg	ggagggcctt	1500
cgtgcgtcgc	cgcgcgcgcg	tccccttctc	catctccagc	ctcggggcgt	ccgcaggggg	1560
acggctgcct	tcggggggga	cggggcaggg	cgggggtcgg	cttctggcgt	gtgacggggc	1620
gctctagagc	ctctgctaac	catgttcatg	ccttcttctt	tttctctacg	ctcctgggca	1680
acgtgctggt	tggtgtgctg	tctcatcatt	ttggcaaaag	attcgccacc	atggtgagca	1740
agggcgagga	gctgttcacc	ggggtggtgc	ccatcctggt	cgagctggac	ggcgacgtaa	1800
acggccacaa	gttcagcgtg	tcgggcgagg	gcgagggcga	tgccacctac	ggcaagctga	1860
ccctgaagtt	catctgcacc	accggcaagc	tgcccggtgc	ctggcccacc	ctcgtgacca	1920
ccctgacct	cggcgtgcag	tgcttcagcc	gctaccccca	ccacatgaag	cagcacgact	1980
tcttcaagtc	cgccatgccc	gaaggctacg	tccaggagcg	caccatcttc	ttcaaggagc	2040
acggcaacta	caagaccgcg	gccgaggtga	agttcgaggg	cgacaccctg	gtgaaccgca	2100
tcgagctgaa	gggcatcgac	ttcaaggagg	acggcaacat	cctggggcac	aagctggagt	2160
acaactacaa	cagccacaac	gtctatatca	tgcccgacaa	gcagaagaac	ggcatcaagg	2220
tgaacttcaa	gatccgccac	aacatcgagg	acggcagcgt	gcagctcgcc	gaccactacc	2280
agcagaacac	ccccatcggc	gacggccccc	tgctgctgcc	cgacaaccac	tacctgagca	2340
ccagtcgcgc	cctgagcaaa	gaccccaacg	agaagcgcg	tcacatgggt	ctgctggagt	2400
tcgtgaccgc	cgccgggatc	actctcgcca	tggacgagct	gtacaagtaa	gaattcactc	2460
ctcaggtgca	ggctgcctat	cagaagggtg	tggctgggtg	ggccaatgcc	ctggctcaca	2520
aataccactg	agattctttt	ccctctgcca	aaaattatgg	ggacatcatg	aaagcccttg	2580
agcatctgac	ttctggctaa	taaaggaaat	ttattttcat	tgcaatagtg	tggttggatt	2640
ttttgtgtct	ctcactcgga	aggacatagt	ggagggcaaa	tcatttaaaa	catcagaatg	2700
agtatttgg	ttagagtttg	gcaacatgtg	ccatatgctg	gctgccatga	acaaaggtgg	2760
ctataaagag	gtcatcagta	tatgaaacag	ccccctgctg	tccattcctt	attccataga	2820
aaagccttga	cttgaggtta	gatttttttt	atatattgtt	ttgtgttatt	tttttcttta	2880
acatccctaa	aattttctct	acatgtttta	ctagccagat	ttttctcctc	ctcctgacta	2940
ctcccagtca	tagctgtccc	tcttctctta	tgaagatccc	tcgacctgca	gcccagcttt	3000
ggcgtaatca	tggtcatagc	tggttctcgt	gtgaaattgt	tatccgctca	caattccaca	3060
caacatacga	gccggaagca	taaagtgtaa	agcctggggg	gcctaagtga	tgagctaact	3120

-93-

cacattaatt	gcgttgccgt	cactgcccgc	tttccagtcg	ggaaacctgt	cgtgccagcg	3180
gatccgcac	tcaattagtc	agcaaccata	gtcccgcgcc	taactccgcc	catccccgcc	3240
ctaactccgc	ccagttccgc	ccattctccg	ccccatggct	gactaatttt	ttttatttat	3300
gcagaggccg	aggccgcctc	ggcctctgag	ctattccaga	agtagtgagg	aggctttttt	3360
ggaggctagt	ggatcccccg	ccccgtatcc	cccaggtgtc	tgcaggtcca	aagagcagcg	3420
agaagcggtc	agaggaaaag	gatcccgtgc	caccttcccc	gtgcccgggc	tgtccccgca	3480
cgctgccggc	tccgggatgc	ggggggagcg	ccggaccgga	gcggagcccc	gggaggctcg	3540
ctgctgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	ggggggggct	3600
gtccccgtga	gcggatccgc	ggccccgtat	ccccaggtg	tctgcaggct	caaagagcag	3660
cgagaagcgt	tcagaggaaa	gcgatcccg	gccaccttcc	ccgtgcccg	gctgtccccg	3720
cacgctgccg	gctcggggat	gcggggggag	gcggggaccg	gagcggagcc	ccggggggct	3780
cgctgctgcc	ccctagcggg	ggaggggacgt	aattacatcc	ctggggggctt	tggggggggg	3840
ctgtccccgt	gagcggatcc	gcggccccgt	atccccagg	tgtctgcagg	ctcaaagagc	3900
agcgagaagc	gttcagagga	aagcgatccc	gtgccacctt	ccccgtgccc	gggctgtccc	3960
cgacagctgc	cggtcggggg	atgcgggggg	agcgcgggac	cgagcgggag	ccccggggcg	4020
ctcgctgctg	ccccctagcg	ggggaggggac	gtaattacat	ccctgggggg	tttggggggg	4080
ggctgtcccc	gtgacgggat	ccgcggcccc	gtatccccca	gggtgtctga	ggctcaaaga	4140
gcagcgagaa	gcgttcagag	gaaagcgatc	ccgtgccacc	ttccccgtgc	ccgggtgtgc	4200
cccgcacgct	gccggctcgg	ggatgcgggg	ggagcggccg	accggagcgg	agccccgggc	4260
ggctcgctgc	tgccccctag	cgggggaggg	acgttaattac	atccctgggg	gctttggggg	4320
ggggctgtcc	ccgtgagcgg	atccgcggcc	ccgtatcccc	cagggtgtctg	caggctcaaa	4380
gagcagcgag	aagcgttcag	aggaaaagcg	tcccgtgcca	ccttccccgt	gccccgggctg	4440
tccccgcacg	ctgcggctc	ggggatgcgg	ggggagcgcc	ggaccggagc	ggagccccgg	4500
gcggctcgct	gctgccccct	agcgggggag	ggacgttaatt	acatccctgg	gggctttggg	4560
gggggggctg	ccccgtgagc	ggatccgcgg	ccccgtatcc	cccaggtgtc	tgcaggctca	4620
aagagcagcg	agaagcgttc	agaggaaaag	gatcccgtgc	caccttcccc	gtgcccgggc	4680
tgtccccgca	cgctgccggc	tccgggatgc	ggggggagcg	ccggaccgga	gcggagcccc	4740
ggggcggtcg	ctgctgcccc	ctagcggggg	agggagcgtaa	ttacatccct	gggggctttg	4800
ggggggggct	gtccccgtga	gcggatccgc	ggggctgcag	gaattcgatt	gaagcctgct	4860
ttttttatact	aacttgagcg	aaatcaagct	cctaggctttt	tgcaaaaagc	taacttggtt	4920
attgcagctt	ataatgggta	caaataaagc	aatagcatca	caaatttcac	aaataaagca	4980
tttttttcac	tgcattctag	ttgtggtttg	tccaaactca	tcaatgtatc	ttatcatgtc	5040
tggatccgct	gcattaatga	atcgcccaac	gcgcggggag	aggcgggttg	cgtatttggg	5100
gctcttcocg	ttcctcgtc	actgactcgc	tgcgctcgg	cgttcggctg	cggcgagcgg	5160
tatcagctca	ctcaaaggcg	gtaatacgg	tatccacaga	atcaggggat	aacgcaggaa	5220
agaacatgtg	agcaaaaagg	cagcaaaaag	ccaggaaccg	taaaaaggcc	gcgttgctgg	5280
cgtttttcca	taggtccgc	ccccctgacg	agcatcacaa	aatcgcagc	tcaagtccag	5340
ggtagcgaaa	cccgcacagg	ctataaagat	accaggcggt	tccccctgga	agctccctcg	5400
tgcgctctcc	tggtccgacc	ctgcccgtta	ccggatacct	gtccgccttt	ctcccttcg	5460
gaagcggtgg	gctttctcaa	tgctcacgtc	tgaggtatct	cagttcgggtg	taggtcggtc	5520
gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	5580
gtaactatcg	tcttgagtcc	aaccgggtaa	gacacgactt	atcgccactg	gcagcagcca	5640
ctggtaaacg	gttcagcaga	gcgagggtatg	taggcgggtgc	tacagagtgc	ttgaagtgg	5700
ggcctaacta	cggctacact	agaaggacag	tatttggtat	ctgcgctctg	ctgaagccag	5760
ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	5820
gtgggtttttt	tggttgcaag	cagcagatta	ccgcagagaaa	aaaaggatct	caagaagatc	5880
ctttgatctt	ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	taagggattt	5940
tggatcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	aatgaagtt	6000
ttaaatcaat	ctaaagtata	tatgagtata	cttggtctga	cagttaccaa	tgcttaatca	6060
gtgaggcacc	tatctcagcg	atctgtctat	ttcggtcatc	catagtggcc	tgactccccg	6120
tcgtgtagat	aactacgata	cgggagggct	taccatctgg	ccccagtgct	gcaatgatac	6180
cgcgagaccc	acgctcacgg	gctccagatt	tatcagcaat	aaaccagcca	gccggaagg	6240
ccgagcgag	aagtggctct	gcaactttat	ccgctcccat	ccagtctatt	aattgttgcc	6300
gggaagctag	agtaagtatg	tcgccagtta	atagtttgcg	caacgttgtt	gccattgcta	6360
caggcatcgt	gggtgtcagc	tcgtcgtttg	gtatggcttc	attcagctcc	ggttcccaac	6420
gatcaaggcg	agttacatga	tcccccatgt	tgtgcaaaaa	agcggtagc	tccttcggtc	6480
ctccgatcgt	tgtcagaagt	aagttggccg	caggtttatc	actcatgggt	atggcagcac	6540
tgcaaatctc	tcttactgtc	atgccatccg	ttaagtgcct	ttctgtgact	ggtagtact	6600
caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgtaa	6660
tacgggataa	taccgcgcca	catagcagaa	ctttaaaggt	gctcatcatt	ggaaaacgtt	6720
cttcggggcg	aaaactctca	aggatctttc	cgctgttgag	atccagttcg	atgtaaccca	6780
ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	6840
aaacaggga	gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	tggtgaatac	6900
ctacactctt	cctttttcaa	tattattgaa	gcatttatca	gggttatgtg	ctcatgagcg	6960
gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	7020
gaaaagtgcc	acctggtcga	cggtatcgat	aagcttgata	tcgaattcct	gcagccccgc	7080
ggatccgctc	acggggagac	cccccccca	aagccccag	ggatgtaatt	acgtccctcc	7140

-94-

```
cccgcctaggg  ggcagcagcg  agccgcccgg  ggctccgctc  cggctccggcg  ctccccccgc  7200
atccccgagc  cggcagcgtg  cggggacagc  ccgggcacgg  ggaaggtggc  acgggatcgc  7260
tttccctctga  acgcttctcg  ctgctctttg  agcctgcaga  cacctggggg  atacgggggc  7320
gcggatccgc  tcacggggac  agcccccccc  caaagccccc  agggatgtaa  ttacgtccct  7380
ccccgcctag  ggggcagcag  cgagccgccc  ggggctccgc  tccggctccgg  cgctcccccc  7440
gcacccccga  gccggcagcg  tgcggggaca  gcccgggcac  ggggaaggtg  gcacgggatc  7500
gctttctctt  gaacgcttct  cgctgctctt  tgagcctgca  gacacctggg  ggatacgggg  7560
ccgcgggatcc  gctcacgggg  acagcccccc  cccaaagccc  ccagggatgt  aattacgtcc  7620
ctcccccgct  agggggcagc  agcgagccgc  ccggggctcc  gctccgggtc  ggcgtcccc  7680
ccgcaccccc  gagccggcag  cgtgccccga  cagcccgggc  acggggaagg  tggcacggga  7740
tcgctttctt  ctgaacgctt  ctgctgctc  tttgagcctg  cagacacctg  ggggatacgg  7800
ggccgcggat  ccgctcacgg  ggacagcccc  ccggggctcc  cccagggat  gtaattacgt  7860
ccctcccccg  ctagggggca  gcagcgagcc  gcccggggct  ccgctccggt  ccggcgctcc  7920
ccccgcaccc  ccgagccggc  agcgtgcggg  gacagccggg  gcacggggaa  ggtggcacgg  7980
gatcgctttt  ctctgaacgc  ttctcgctgc  tctttgagcc  tgcagacacc  tgggggatac  8040
ggggcccgcg  atccgctcac  ggggacagcc  ccccccaaaa  gcccccaggg  atgtaattac  8100
gtccctcccc  cgctaggggg  cagcagcgag  ccgccccggg  ctccgctccg  gtcggcgctc  8160
cccccgcat  ccccgagccg  gcagcgctgc  gggacagccc  gggcacgggg  aaggtggcac  8220
gggatcgctt  tcctctgaac  gcttctcgct  gctctttgag  cctgcagaca  cctgggggat  8280
acggggccgc  ggatccgctc  acggggacag  ccccccccca  aagccccag  ggatgtaatt  8340
acgtccctcc  ccgctaggg  ggcagcagcg  agccgcccgg  ggctccgctc  cggcccgcg  8400
ctccccccgc  atccccgagc  cggcagcgtg  ccgggcacgg  ggaaggtggc  cacctggggg  8460
acgggatcgc  tttcctctga  acgcttctcg  ctgctctttg  agcctgcaga  8520
a  8521
```

&lt;210&gt; 124

&lt;211&gt; 8851

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; p18EPOcDNA Plasmid

&lt;400&gt; 124

```
cagttgcccgg  cccgggtcggc  cagggcggaac  tcccgccccc  acggctgctc  gccgatctcg  60
gtcatggccgg  gcccgagggc  gtcccggaag  ttcgtggaca  cgacctccga  ccactcggcg  120
tacagctcgt  ccaggccggc  caccacacac  caggccaggg  tggtgtccgg  caccacctgg  180
tcctggaccg  cgetgatgaa  caggggtcacg  tcgtcccggg  ccacaccggc  gaagtgcgtc  240
tccacgaagt  cccgggagaa  cccgagccgg  tccgtccaga  actcgaccgc  tccggcgacg  300
tcgcgcggcg  tgagcacggg  aacggcactg  gtcaacttgg  ccatggatcc  agatttcgct  360
caagttagta  taaaaaagca  ggcttcaatc  ctgcagagaa  gcttgatata  gaattcctgc  420
agcccccgcg  atccgctcac  ggggacagcc  ccccccaaaa  gcccccaggg  atgtaattac  480
gtccctcccc  cgctaggggg  cagcagcgag  ccgcccgggg  ctccgctccg  gtcggcgctc  540
ccccccgcac  ccccgagccg  gcagcgctgc  gggacagccc  gggcacgggg  aaggtggcac  600
gggatcgctt  tcctctgaac  gcttctcgct  gctctttgag  cctgcagaca  cctgggggat  660
acggggccgc  ggatccgctc  acggggacag  ccccccccca  aagccccag  ggatgtaatt  720
acgtccctcc  ccgctaggg  ggcagcagcg  agccgcccgg  ggctccgctc  cggcccgcg  780
ctccccccgc  atccccgagc  cggcagcgtg  ccgggcacgg  ggaaggtggc  cacctggggg  840
acgggatcgc  tttcctctga  acgcttctcg  ctgctctttg  agcctgcaga  cactcggggg  900
atacgggggc  gcggatccgc  tcacggggac  agcccccccc  caaagccccc  agggatgtaa  960
ttacgtccct  cccccgctag  ggggcagcag  cgagccgccc  ggggctccgc  tccggctccg  1020
cgctcccccc  gcacccccga  gccggcagcg  tgcggggaca  gcccgggcac  ggggaaggtg  1080
gcacgggatc  gctttctctt  gaacgcttct  cgctgctctt  tgagcctgca  gacacctggg  1140
ggatacgggg  ccgcggtacc  gctcacgggg  acagcccccc  cccaaagccc  ccagggatgt  1200
aattacgtcc  ctcccccgct  agggggcagc  agcgagccgc  ccggggctcc  gctccgggtc  1260
ggcgctcccc  ccgcaccccc  gagccggcag  cgtgccccga  cagcccgggc  acggggaagg  1320
tggcacggga  tgcctttcct  ctgaacgctt  ctgctgctc  tttgagcctg  cagacacctg  1380
ggggatacgg  ggcgcgggat  gccgctcagc  ctcagacccc  cccccaaaag  cccacagggg  1440
gtaattacgt  cctcccccg  ctagggggca  gcagcgagcc  gcccggggct  ccgctccggg  1500
ccggcgctcc  ccccgcatcc  ccgagccggc  agcgtgcccc  gacagccggg  gcacggggaa  1560
ggtggcacgg  gatcgcttcc  ctctgaacgc  tcttctgctc  tctttgagcc  tgcagacacc  1620
tgggggatac  ggggcccggg  atccgctcac  cgctaggggg  ccccccacaa  gcccccaggg  1680
atgtaattac  gtccctcccc  ccccgagccg  gccagcgag  ccgcccgggg  ctccgctccg  1740
gtccggcgct  aaggtggcac  cctgggggat  cctgggggat  gggcacgggg  gggcacgggg  1800
cctgggggat  tagttcatag  cccatatatg  gagttccggc  ttacataact  tacggtaaat  1860
ggcccgcccg  1920
ggcccgcccg  1980
```

gctgaccgcc	caacgacccc	cgcccattga	cgtcaataat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gacttttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgcccact	2100
tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagtaga	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcata	gctattacca	tgggtcgagg	tgagcccac	gttctgtctc	2280
actctcccca	tctccccccc	ctccccaccc	ccaattttgt	atltatltat	tttttaatta	2340
ttttgtgcag	cgatgggggc	gggggggggg	ggggcgcgcg	ccaggcgggg	cgggggcgggg	2400
cgagggggcg	ggcgggggcg	ggcgagagag	tgcgggcgca	gccaatcaga	gcggcgcgct	2460
cgaaaagt	ctttttatgg	cgaggcgggc	gcggcgggcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtgcgtg	cgttgccttc	gccccgtgcc	ccgctccgcg	ccgctccgcg	2580
ccgcccgcgc	cggtctgac	tgaccgcgtt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctctcccg	ggctgtaatt	agcgcttgg	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
gcgtgaaagc	cttaaagggc	tccgggaggg	ccctttgtgc	gggggggagc	ggctcggggg	2760
gtgcgtgcgt	gtgtgtgtgc	gtggggagcg	cccgcgctgc	cccgcgctgc	ccggcggtgc	2820
tgagcgtgcg	ggggcgggcg	cgggcgcttt	tgcgctccgc	gtgtgcgcga	ggggagcgcg	2880
gcccggggcg	gtgccccgcg	gtgcgggggg	gctgcgaggg	gaacaaaggc	tgcggtcggg	2940
gtgtgtgcgt	gggggggtga	gcaggggggt	tggggcgggc	ggtcgggctg	taaccccccc	3000
ctgcaacccc	ctccccag	tgctgagcac	ggcccggtt	cggtgtcggg	gctccgtgcg	3060
gggcgtggcg	cggggctcgc	cgtgccgggc	gggggggtgg	ggcaggtggg	ggtgccgggc	3120
ggggcggggc	cgccctcggg	cggggagggc	tccggggagg	ggcgcgcgcg	ccccggagcg	3180
ccggcggtgc	tcgagcgcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
gggcgcaggg	acttcctttg	tcccaaactc	ggcgagagcg	aaatctggga	ggcgccggcg	3300
cacccctct	agcggggcg	ggcgaagcgg	tgccggcgcc	gcaggaagga	aatggggcg	3360
gagggccttc	gtgcgtcgcc	gcgcgcgggt	cccccttccc	atctccagcc	tcggggctgc	3420
cgagggggga	cggtgcctt	cgggggggac	ggggcgaggg	ggggttcggc	ttctggcggt	3480
tgaccggcg	ctctagaatg	gggggtgcacg	aatgtcctgc	ctggctgtgg	cttctcctgt	3540
ccctgctgtc	ggcctcctcg	ggcctcccag	tcctgggcgc	cccaccacgc	ctcatctgtg	3600
acagccgagt	cctggagagg	tacctcttgg	aggccaagga	ggccgagaat	atcacgacgg	3660
gctgtgctga	acactgcagc	ttgaatgaga	atatcactgt	cccagacacc	aaagttaatt	3720
tcctatgcctg	gaagaggatg	gaggtcgggc	agcaggccgt	agaagtctgg	cagggcctgg	3780
ccctgctgtc	ggaagctgtc	ctgcggggcc	aggccctgtt	ggtcaactct	tcccagccgt	3840
gggagccctt	gcagctgcat	gtggataaag	ccgtcagtgg	ccttcgcagc	ctcaccactc	3900
tgcttcgggc	tcgtggagcc	cagaaggaag	tccatgtccc	tccagatgcg	gcctcagttg	3960
ctccactccg	aacaatcact	gctgacactt	tccgcaaact	cttcagagtc	tactccaatt	4020
tcctccgggg	aaagctgaag	ctgtacacag	gggaggccct	caggacaggg	gacagatgac	4080
gtatacgtaa	gaattcactc	ctcaggtgca	agcactgcct	cagaagggtg	tggctgtgtg	4140
ggccaatgcc	ctggctcaca	aataccactg	agatcttttt	ccctctgcca	aaaattatgg	4200
ggacatcatg	aagccccttg	agcatctgac	ttctggctaa	taaaggaaat	ttattttcat	4260
tgcaatagtg	gtttggaatt	ttttgtgtct	ctcactcgga	aggacatatg	ggaggcgaaa	4320
tcatttaaaa	catcagaatg	agtatttgg	ttagagtgtg	gcaacatatg	ccatatgctg	4380
gctgccatga	acaaagggtg	ctataaaagag	gtcatcagta	tatgaaacag	ccccctgtgt	4440
tccattcctt	attccataga	aaagccttga	cttgaggtta	gatttttttt	atatttttgt	4500
ttgtgttatt	tttttcttta	acatccctaa	aatttttctt	acatgtttta	ctagccagat	4560
ttttcctcct	ctcctgacta	ctcccagtc	tagctgtccc	tcttctctta	tgaagatccc	4620
tcgactgcga	gccccagctt	gcatgcctgc	aggtgcactc	tagtggatcc	ccgcctccgt	4680
atcccccgag	tgctctgagg	ctcaaagagc	agcgagaagc	gttcagagga	aagcgatccc	4740
gtgccacctt	ccccgtgccc	gggctgtccc	cgacagctgc	cggctcgggg	atgcgggggg	4800
agcgccggag	cggagcgagg	ccccggggcg	ctcgctgctg	ccccctagcg	ggggagggag	4860
gtaattacat	ccctgggggg	tttggggggg	ggctgtcccc	gtgagcggat	ccgcggcccc	4920
gtatccccca	gggtgtctga	ggctcaaaga	gcagcgagaa	gcgttcagag	gaaagcgatc	4980
ccgtgcccac	ttccccgtgc	ccgggctgtc	cccgcacgct	gcccggctcg	ggatgcgggg	5040
ggagcgccgg	accggagcgg	agccccgggc	ggctcgtgc	tgccccctag	cgggggaggg	5100
acgtaattac	atccctgggg	gctttggggg	ggggctgtcc	ccgtgagcgg	atccgcggcc	5160
ccgtatcccc	caggtgtctg	caggctcaaa	gagcagcgag	aagcgttcag	aggaaagcga	5220
tccgtgcca	ccttccccgt	gcccgggctg	tccccgcacg	ctgccggctc	ggggatgcgg	5280
gggagcgcc	ggaccggagc	ggagcccccg	gcggctcgct	gctgccccct	agcgggggag	5340
ggagcgtaat	acatccctgg	gggctttggg	ggggggctgt	ccccgtgagc	ggatcccgcg	5400
ccccgtatcc	cccaggtgtc	tgccaggtca	aagagcagcg	agaagcgttc	agaggaaagc	5460
gatcccgctg	caccttcccc	gtgccccggc	tgtccccgca	cgctgcccgc	tcggggatgc	5520
gggggggagc	ccggaccgga	gcccagcccc	ggcggtcgct	ctgctgcccc	ctagcggggg	5580
agggagcgt	ttacatccct	ggggggcttt	ggggggggct	gtccccctga	gcggatccgc	5640
ggccccgtat	cccccggtg	tctgcaggct	caaagagcag	cgagaagcgt	tcagaggaaa	5700
gcgatcccg	gccaccttcc	ccgtgccccg	gctgtccccg	cacgctgccc	gctcggggat	5760
gcggggggag	cgccggaccg	gagcggagcc	ccggggcgct	cgctgctgcc	ccctagcggg	5820
ggagggagct	aattacatcc	ctggggggct	tggggggggg	ctgtccccgt	gagcggatcc	5880
gcggccccgt	atcccccagg	tgtctgcagg	ctcaaagagc	agcgagaagc	gttcagagga	5940
aagcgatccc	gtgccacctt	ccccgtgccc	gggctgtccc	cgcacgctgc	cggctcgggg	6000

-96-

```

atgcggggggg agcgccggac cggagcggag cccggggcgg ctcgctgctg cccctagcgg 6060
ggggaggggac gtaattacat ccctgggggg tttggggggg ggctgtcccc gtgagcggat 6120
ccgcgggggct gcagggaattc gtaatcatgg tcatagctgt ttctgtgtgt aaattgttat 6180
ccgctcacaa ttccacacaa catacagagcc ggaagcataa agtgtaaagc ctgggggtgcc 6240
taatgagtga gctaactcac attaatgctg ttgcgctcac tgcccgcttt ccagtcggga 6300
aacctgtcgt gccagctgca ttaatgaatc ggccaacgcy cggggagagg cggtttgctg 6360
attgggcgct ctcccgcttc ctgcctcact gactcgctgc gctcggctcgt tcggctgcgg 6420
cgagcgggat cagctcactc aaaggcggta atacggttat ccacagaatc aggggataac 6480
gcaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcy 6540
ttgctggcgt ttttccatag gctccgcccc cctgacgagc atcacaaaaa tcgacgctca 6600
agtcagaggt ggcgaacccc gacaggacta taaagatacc aggcgtttcc ccctggaagc 6660
tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg gatacctgtc cgcctttctc 6720
ccttcgggaa gcgtggcgct ttctcatagc tcacgctgta ggtatctcag ttcggtgtag 6780
gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg tttagcccca ccgctgcgcc 6840
ttatccggtt tgagtccaac cgggtaagac acgacttatc ggcactggca 6900
gcagccactg gtaacaggat tagcagagcg aggtatgtag gcggtgctac agagtctctg 6960
aagtgggtggc ctaactacgg ctacactaga aggcacgtat ttggtatctg cgctctgcgt 7020
aagccagtta ccttcggaaa aagagtgtgt agtcttgat ccggcaaaaca aaccacgctg 7080
ggtagcgggtg gtttttttgt ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa 7140
gaagatcctt tgatcttttc tacgggtctc gacgctcagt ggaacgaaaa ctcacgttaa 7200
gggatttttg tcatgagatt atcaaaaagg atcttcacct agatcctttt aaattaaaaa 7260
tgaagtttta aatcaatcta aagtatatat gagtaaactt ggtctgacag ttaccaatgc 7320
ttaatcagtg aggcacctat ctacgcgatc tgtctatttc gtatcatccat agttgcctga 7380
ctccccgtcg ttagataaac tacgatacgg gagggcttac catctggccc cagtgtgca 7440
atgataccgc gagaccacag ctacccggct ccagatttat cagcaataaa ccagccagcc 7500
ggaagggcgg agcgcagaag tggctcctgca actttatccg cctccatcca gtctattaat 7560
tggtgcgggg aagctagagt aagtagttcg ccagttaata gtttgcgcaa cgttgttgcc 7620
attgctacag gcatcggtgt gtcacgctcg tctgttggtg tggttccatt cagctccggg 7680
tcccacagat caaggcgagt ccatgtatcc cccatgttgt gcaaaaaagc ggttagctcc 7740
ttcggtcctc ccatcggtgt cagaagttaa tggccgcag tgttatcact catggttatg 7800
gcagcactgc ataattctct tactgtcatg ccatccgtaa gatgcttttc tgtgactggg 7860
gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc gaccgagttg ctcttgcccg 7920
gcgtcaatac ggggcaatag cgcgccacat agcagaactt taaaagtgtc catcattgga 7980
aaacgttctt cggggcgaaa actctcaagg atcttaccgc tggtgagatc cagttcgatg 8040
taaccacact gtgcacccaa ctgatcttca gcatctttta ctttcaccag cgtttctggg 8100
tgagcaaaaa caggaaggca aaatgccgca ttttcaatat taagggcgac acggaaaatg 8160
tgaatactca tactcttcct ttttcaatat tattgaaaga tttatcaggg 8220
atgagcggat acatatttga atgtattttag aaaaaataaac aaataggggt tccgcgcaca 8280
ttccccgaa aagtgccacc tgacgtagtt ggcgtcgcaa agcccgccga agcgggcttt 8340
attaccaagc gaagcgccat tgcctattca ggtcgcgcaa ctgttgggaa gggcgatcgg 8400
tgccggcctc ttcgctatta cgccagctgg cgaagggggg atgtgctgca aggcgattaa 8460
gttgggtaac gccagggttt tcccagtcac gacgttgtaa aacgacggcc agtcgtaat 8520
acgactcact actaggggtg ctagaggggt cagcgggtata cagacatgat aagatacatt 8580
gatgagtttg gacaaaccac aactagaatg cagtgaaaaa aatgctttat ttgtgaaatt 8640
tgtgatgcta ttgctttatt tgtaaccatt ataagctgca ataaacaagt tgggggtggg 8700
gaagaactcc agcatgagat ccccgcgctg gaggtcatc cagccggcgt cccggaaaac 8760
gattccgaag cccaaccttt catagaaggc ggcgtggaa tcgaaatctc gtagcacgtg 8820
tcagtcctgc tcctcgcca cgaagtgcac g g 8851

```

&lt;210&gt; 125

&lt;211&gt; 10474

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; p18genEPO Plasmid

&lt;400&gt; 125

```

cagttgccgg ccgggtcgcg cagggcgaaac tcccggcccc acggtctgctc gccgatctcg 60
gtcatggccg gccggaggcg gtcccggaaac ttcggtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccgcg caccacacagg caggccaggg tgttgtccgg caccactggt 180
tcctggaccg cgctgatgaa cagggtcacg tctcccggga ccacaccggc gaagtctgct 240
tccacgaagt cccgggagaa cccgagccgg tcggtccaga actcgaccgc tccggcgacg 300
tcgcgcggcg tgagcaccgg aacgycactg gtcacttggg ccattggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgatata gaattcctgc 420
agcccccgcg atccgctcac ggggacagcc ccccccaaaa gcccccaggg atgtaattac 480
gtccctcccc cgtagggggg cagcagcgag ccgcccgggg ctcgctcccg gtcggcgctc 540

```



ccccccgcat	ccccgagccg	gcagcgtgcg	gggacagccc	gggacagggg	aaggtggcac	600
gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	cctgggggat	660
acggggccgc	ggatccgctc	acggggacag	cccccccca	aagccccag	ggatgtaatt	720
acgtccctcc	cccgctaggg	ggcagcagcg	agccgcccgg	ggctccgctc	cggtccggcg	780
ctccccccgc	atccccgagc	cggcagcgtg	cggggacagc	ccgggacagg	ggaaggtggc	840
acgggatcgc	tttctctga	acgttctcgc	ctgctctttg	agcctgcaga	cacctggggg	900
atacggggcc	gcggatccgc	tcacggggac	agcccccccc	caaagccccc	agggatgtaa	960
ttacgtccct	cccccgctag	ggggcagcag	cgagccgccc	ggggctccgc	tcgggtccgg	1020
cgtccccccc	gcacccccga	gcccgcagcg	tgccggggaca	gcccggggcac	gggggaaggtg	1080
gcacggggatc	gctttctctc	gaacgcttct	cgtctctctt	tgagcctgca	gacacctggg	1140
ggatacggggg	cccgcggtacc	gctcacgggg	acagcccccc	cccaaagccc	ccaggggatgt	1200
aattacgtcc	ctcccccgct	agggggcagc	agcagcgccg	ccggggctcc	gctccggctcc	1260
ggcgctcccc	ccgcatcccc	gagccggcag	cgtgcggggg	cagcccgggc	acgggggaagg	1320
tggcacggga	tcgctttctc	ctgaacgctt	ctcgctgctc	tttgagcctg	cagacacctg	1380
ggggatcggg	ggcccgcgat	ccgctcacgg	ggacagcccc	ccccaaagc	ccccgggat	1440
gtaattacgt	ccctcccccg	ctagggggca	gcagcgagcc	gcccgggggt	ccgctccggg	1500
ccggcgctcc	ccccgcctcc	ccgagccggc	agcgtgcggg	gacagcccg	gcacggggaa	1560
gggtggcagg	gatcgcttcc	ctctgaacgc	ttctcgctgc	tccttgagcc	tgacagaccc	1620
tgggggatac	ggggcccgcg	atccgctcac	ggggacagcc	ccccccaaa	gccccagggg	1680
atgtaattac	gtccctcccc	cgttaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	1740
gtccggcgct	cccccgcat	ccccgagccg	gcagcgtgcg	gggacagccc	gggacagggg	1800
aaggtggcac	gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttattaata	gtaatcaatt	acggggctcat	1920
tagttcatag	cccatatatg	gagttccgcg	ttacataact	tacggtaaat	ggcccgcgctg	1980
gctgaccgcc	caacgacccc	gcgccattga	cgtcaataat	gacgtatggt	cccatagtaa	2040
cgccaatagg	gactttccat	tgaogtcaat	gggtggacta	tttacggtaa	actgcccact	2100
tggcagtaga	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagtaga	tgacctatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcate	gctattacca	tggggtcgagg	tgagccccac	gttctgcttc	2280
actctcccc	tctccccccc	ctccccccc	ccaattttgt	atttatttat	tttttaatta	2340
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cggggcgggg	2400
cgagggggcg	ggcgggggcg	ggcgggagag	tgccggcgga	gccaatcaga	gcggcgcgct	2460
ccgaaagtgt	ccttttatgg	cgaggcgggc	gcggcgggcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtgcgtg	cgttgccctc	gccccgtgcc	ccgctccgcg	ccgcctcgcg	2580
ccgcccgcgc	cggtcttgac	tgaccgcgtt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctctcccg	ggctgtaatt	agcgcttggt	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
gcgtgaaagg	cttaaagggg	tcggggaggg	ccctttgtgc	ggggggggag	ggctcggggg	2760
gtgctgtgct	gtgtgtgtgc	gtggggagcg	ccgcgtgagg	cccgcgctgc	ccggcgggctg	2820
tgagcgctgc	gggcgcggcg	cggggctttg	tgcgctccgc	gtgtgcgcga	ggggagcgcg	2880
gcccggggcg	gtgccccgcg	gtgcgggggg	gctgcgaggg	gaacaaaggc	tgctgtcggg	2940
gtgtgtgctg	gggggggtga	gcaggggggtg	tgggcgcggc	ggtcgggctg	taaccccccc	3000
ctgcaccccc	ctccccgagt	tgctgagcac	ggcccggctt	cgggtgcggg	gctccgtgcg	3060
gggcgtggcg	cggggctcgc	cgtgcccggc	gggggggtggc	ggcaggtggg	gggtgcgggc	3120
ggggcggggc	cgcctcgggc	cggggagggc	tcggggggagg	ggcgcgggcg	ccccggagcg	3180
ccggcgggctg	tcgaggcgcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
gggcgcaggg	acttcccttg	tcctaaatct	ggcgagcccg	aaatctggga	ggcgccggcg	3300
cacccccctc	agcgggcgcg	ggcgaaagcg	tgccggcgccg	gcaggaagga	aatgggcggg	3360
gagggccttc	gtgctgcgcc	gcgcgcgcgt	cccccttctc	atctccagcc	tcgggggctgc	3420
cgcaggggga	cggctgcctt	cggggggggc	ggggcagggc	gggggtcggc	ttctggcggt	3480
tgaccggcg	ctctagatgc	atgctcgagc	ggccgcaggt	gtgatggata	tcctgcagaat	3540
tcgccccttc	tagaatgggg	gtgcacgggtg	agtactcgcg	ggctggggcg	ccccgcccgc	3600
ccgggtccct	gtttgagcgg	ggatttagcg	ccccggctat	tgccagggag	gtggctgggt	3660
tcaaggaccg	gcgacttgct	aaggaccccg	gaagggggag	gggggtgggg	tgccctccacg	3720
tgccagcggg	gacttggggg	agtccttggg	gatggcaaaa	acctgacctg	tgaaggggac	3780
acagtttggg	ggttgagggg	aagaagggtt	gggggttctg	ctgtgccagt	ggagagggaag	3840
ctgataagct	gataaacctg	gcgctggagc	caccacttat	ctgccagagg	ggaagcctct	3900
gtcacaccag	gattgaagtt	tggccgggag	agtggatgct	ggtagctggg	gggtgggggt	3960
gcacacggca	gcaggattga	atgaaggcca	gggaggcagc	acctgagtg	ttgcatgggt	4020
ggggacagga	aggacgagct	ggggcagaga	cgtggggatg	aaggaaagctg	tccttcacaca	4080
gccacccttc	tcctccccgc	cctgactctc	agcctggcta	tctgttctag	aatgtcctgc	4140
ctggctgtgg	cttctcctgt	ccctgctgtc	gctccctctg	ggcctccagc	tcctggggcg	4200
cccaccacgc	ctcatctgtg	acagccgagt	cctggagagg	tacctcttgg	aggccaaggga	4260
ggccgagaa	atcacggtga	gaccccttcc	ccagcacatt	ccacagaact	cacgctcagg	4320
gcttcaggga	actcctccca	gatccaggaa	cctggcactt	gggttggggg	ggagtgggga	4380
agctagacac	tgcccccccta	cataagaata	agtctgggtg	ccccaaacca	tacctggaaa	4440
ctaggcaagg	agcaaaagcca	gcagatccta	cggcctgtgg	gccagggcca	gagccttcag	4500
ggacccttga	ctccccgggc	tgtgtgcatt	tcagacgggc	tgtgctgaac	actgcagctt	4560

gaatgagaat	atcactgtcc	cagacaccaa	agttaatttc	tatgcctgga	agaggatgga	4620
ggtgagttcc	tttttttttt	tttttccttt	cttttggaga	atctcatttg	cgagcctgat	4680
tttggatgaa	agggagaaatg	atcgagggaa	aggtaaaatg	gagcagcaga	gatgaggctg	4740
cctgggcgca	gaggctcacg	tctataatcc	caggctgaga	tggccgagat	gggagaattg	4800
cttgagccct	ggagtttcag	accaacctag	gcagcatagt	gagatcccc	atctctacaa	4860
acatttaaaa	aaattagtca	ggtgaagtgg	tgcagtgtgg	tagtcccaga	tatttggaag	4920
gctgaggcgg	gaggatcgct	tgagcccagg	aatttgaggc	tgcaagtggc	tgtgatcaca	4980
ccactgcact	ccagcctcag	tgacagagt	aggccctgtc	tcaaaaaaga	aaagaaaaaa	5040
gaaaaataat	gagggctgta	tggaatacat	tcattattca	ttcactcact	cactcactca	5100
ttcattcatt	cattcattca	acaagtctta	ttgcatacct	tctgtttgct	cagcttggtg	5160
cttggggctg	ctgaggggca	ggagggagag	ggtgacatgg	gtcagctgac	tcccagagtc	5220
cactccctgt	aggctcgggca	gcaggccgta	gaagtctggc	agggcctggc	cctgctgtcg	5280
gaagctgtcc	tgcggggcca	ggccctgttg	gtcaactcct	cccagccgtg	ggagcccctg	5340
cagctgcatg	tggataaagc	cgtcagtggc	cttcgcagcc	tcaccactct	gcttcgggct	5400
ctgggagccc	aggtgagtag	gagcggacac	ttctgcttgc	cctttctgta	agaaggggag	5460
aagggctctg	ctaaggagta	caggaactgt	ccgtattcct	tccctttctg	tggcactgca	5520
gcgacctcct	gttttctcct	tggcagaagg	aagccatctc	ccctccagat	gcggcctcag	5580
ctgtccact	ccgaacaatc	actgctgaca	ctttccgcaa	actcttcoga	gtctactcca	5640
atttccctcg	gggaagctg	aagctgtaca	caggggaggc	ctgcaggaca	ggggacagat	5700
gacgtacaag	taagaattca	ctcctcaggt	gcagcctgcc	tatcagaagg	tgggtggctgg	5760
tgtggccaat	ggcctgggctc	acaaatacca	ctgagatcct	tttccctctg	ccaaaaatta	5820
tggggacatc	atgaagcccc	ttgagcatct	gacttctggc	taataaagga	aattttatttt	5880
cattgcaata	gtgtgttgga	attttttgtg	tctctcactc	ggaaggacat	atgggagggc	5940
aaatcattta	aaacatcaga	atgagtattt	ggtttagagt	ttggcaacat	atgccatg	6000
ctggctgcca	tgaacaaagg	tggctataaa	gaggtcatca	gtatatgaaa	cagccccctg	6060
ctgtccattc	cttattccat	agaaaagcct	tgacttgagg	ttagattttt	tttatatttt	6120
gttttgtgtt	atttttttct	ttaacatccc	taaaattttc	cttacctgtt	ttactagcca	6180
gatttttctc	cctctcctga	ctactcccag	tcatagctgt	ccctcttctc	ttatgaagat	6240
ccctcgacct	gcagcccaag	cttgcatgcc	tgcaggctga	ctctagtggg	tcccccgccc	6300
cgtatccccc	aggctctctg	aggctcaaag	agcagcgaga	agcgttcaga	ggaaagcgat	6360
cccgtgccac	cttccccgtg	cccgggctgt	ccccgcacgc	tgcgggctcg	gggatgcggg	6420
gggagcgccg	gaccggagcg	gagccccggg	cggctcgctg	ctgcccccta	gcggggggag	6480
gacgttaatt	catccctggg	ggctttgggg	gggggctgtc	cccgtagagc	gatcccgagg	6540
cccgtatccc	ccagggtgtct	gcaggctcaa	agagcagcga	gaagcgttca	gaggaaagcg	6600
atcccgctgc	accttccccg	tgccccgggt	gtccccgcac	gctgcccggc	cggggatgcg	6660
gggggagcgc	cggaccggag	cggagccccg	ggcgctcgcc	tgtgcccccc	tagcggggga	6720
gggacgtaat	tacatccctg	ggggctttgg	gggggggctg	tccccgtgag	cggatccgcg	6780
gccccgtatc	ccccagggtg	ctgcaggctc	aaagagcagc	gagaagcgtt	cagaggaaag	6840
cgtatcccggt	ccaccttccc	cgtgccccgg	ctgtccccgc	acgctgcccg	ctcggggatg	6900
cgggggggagc	gcccggaccgg	agcggagccc	cggggcgctc	gctgctgccc	cctagcgggg	6960
gagggagcgt	attacatccc	tgggggcttt	gggggggggc	tgtccccctg	agcggatccg	7020
cggccccgta	tccccagggt	gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggga	7080
agcgatcccg	tgccaccttc	cccggtgccc	ggctgtcccc	gcacgctgcc	ggctcgggga	7140
tgcggggggga	gcgcccggacc	ggagcgggag	cccggggcgg	tgcgtgctgc	cccctagcgg	7200
gggaggggacg	taattacatc	cctgggggct	tggggggggg	gctgtccccg	tgagcggatc	7260
cgcggccccg	tatccccagc	gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	7320
aaagcgatcc	cgtgccacct	tccccgtgcc	cgggctgtcc	ccgcacgctg	ccggctcggg	7380
gatgcggggg	gagcgccgga	ccggagcggg	gccccgggcg	gctcgctgct	gccccctagg	7440
gggggagggga	cgttaattaca	tccctggggg	ctttgggggg	gggctgtccc	cgtgagcggg	7500
tcccgggccc	cgtatccccc	agggtgtctg	aggctcaaag	agcagcgaga	agcgttcaga	7560
ggaaagcgat	cccggtgccac	cttccccgtg	cccggtctgt	ccccgcacgc	tgcgggctcg	7620
gggatgcggg	gggagcgccg	gaccggagcg	gagccccggg	cggctcgctg	ctgcccccta	7680
gcggggggagg	gacgtaatta	catccctggg	ggctttgggg	gggggctgtc	cccgtgagcg	7740
gatccgcggg	gctgcaggaa	ttcgtaatca	tggctcatagc	tgtttcctgt	gtgaaattgt	7800
tatccgctca	caattccaca	caacatacga	gcccgaagca	taaagtgtaa	agcctggggg	7860
gcctaattgag	tgagctaact	cacattaatt	gcgttgccgt	cactgcccgc	tttccagtcg	7920
ggaaacctgt	cgtgccagct	gcattaatga	atcggccaaac	gcgcggggag	aggcggtttg	7980
cgtattgggc	ttccttcgcg	ttcctcgctc	tgcgctcggt	tgcgctcggt	cgttcggctg	8040
cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	8100
aacgcaggaa	agaacatgtg	agcaaaaagg	cagcaaaaagg	ccaggaaccg	taaaaaagcc	8160
gcgtttctgg	gtgttttcca	taggttcgcg	ccccctgacg	agcatcacaa	aaatcgagcg	8220
tcaagtcaga	ggtggcgaaa	cccagacagga	ctataaagat	accaggcggt	tccccctgga	8280
agctccctcg	tgcgctctcc	tgttccgacc	ctgcccgtta	ccggatacct	gtccgccttt	8340
ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	cagttcggtg	8400
taggtcggtc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	8460
gccttatccg	gtaactatcg	tcttgagctc	aaccgggtaa	gacacgactt	atcgccactg	8520
gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	taggcgggtc	tacagagttc	8580



-99-

ttgaagtggg	ggcctaacta	cggctacact	agaaggacag	tatttggtat	ctgcgctctg	8640
ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	8700
gctggtagcg	gtgggttttt	tggttgcaag	cagcagatta	cgcgagaaaa	aaaaggatct	8760
caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	8820
taagggattt	tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	8880
aatgaagtt	ttaaatcaat	ctaaagtata	tatgagtaaa	cttggctctga	cagttaccaaa	8940
tgcttaatca	gtgaggcacc	tatctcagcg	atctgtctat	tctgttcac	catagttgcc	9000
tgactccccg	tcgtgtagat	aactacgata	cgggagggct	taccatctgg	cccagtgct	9060
gcaatgatac	cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	9120
gccggaaggg	ccgagcgacg	aagtggctct	gcaactttat	ccgcctccat	ccagttctatt	9180
aattgttgcc	gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	9240
gccattgcta	caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	9300
ggttcccaac	gatcaaggcg	agttacatga	tccccatgt	tgtagcaaaaa	agcgggttagc	9360
tccttcgggtc	ctccgacgtc	tgtagaagt	aagttggccg	cagtgttatc	actcatgggt	9420
atggcagcac	tgcataatct	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	9480
ggtagtagt	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	9540
ccggcgtaaa	tacgggataa	taccgcgcca	catagcagaa	ctttaaagt	gctcatcatt	9600
ggaaaaagtt	cttcggggcg	aaaactctca	aggatcttac	cgctgttgag	atccagttcg	9660
atgtaaccca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	9720
gggtgagcaa	aaacaggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	9780
tgtagaatac	tcatactctt	cttttttcaa	tattattgaa	gcatttatca	gggttattgt	9840
ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	gggtccgcgc	9900
acatttcccc	gaaaagtgcc	acctgacgta	gttaacaaaa	aaaagcccg	cgaagccggc	9960
tttattacca	agcgaagcgc	cattcgccat	tcaggtcggt	caactgttgg	gaagggcgat	10020
cggtgccccg	ctcttcgcta	ttacgccagc	tgccgaaagg	gggatgtgct	gcaaggcgat	10080
taagttgggt	aacgccaggg	ttttcccgat	cacgacgttg	taaaacgacg	gccagtcctg	10140
aatacgcact	acttaaggcc	ttgactagag	ggcgacgggt	atacagacat	gataagatac	10200
attgatgagt	ttggacaaac	cacaactaga	atgcagtgaa	aaaaatgctt	tatttgtgaa	10260
atttgtgatg	ctattgcttt	atttgttaac	attataagct	gcaataaaca	agttgggggtg	10320
ggcgaagaac	tcacgatgac	gatccccgcg	ctggaggatc	atccagccgg	cgtcccgga	10380
aacgattccg	aagcccaacc	tttcatagaa	ggcgccgggtg	gaatcgaaat	ctcgtagcac	10440
gtgtcagtc	tgctcctcgg	ccacgaagtg	cacg			10474

&lt;210&gt; 126

&lt;211&gt; 6119

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; p18attBZeoeGFP Plasmid

&lt;400&gt; 126

cagttgccccg	ccgggtcgcg	cagggcggaac	tccccccccc	acggctgctc	gccgatctcg	60
gtcatggccg	gcccggaggc	gtcccgggaa	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggcccg	caccacacac	caggccaggg	tggtgtccgg	caccacctgg	180
tcctggaccg	cgctgatgaa	cagggctcag	tcgtcccggg	ccacaccggc	gaagtcgctc	240
tcacgaaggt	cccgggagaa	cccgagccgg	tcgggtccaga	actcgaccgc	tcggcgagcg	300
tcgcgcgcgg	tgagcacccg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgggctg	caggtcgagg	420
gatcttcata	agagaagagg	gacagctatg	actgggagta	gtcaggagag	gaggaaaaat	480
ctgggtagta	aaactaagta	ggaaaatttt	agggatgtta	aagaaaaaaa	taacacaaaa	540
caaaatataa	aaaaaatcta	acctcaagtc	aaggcttttc	tatggaataa	ggaatggaca	600
gcaggggggt	gtttcatata	ctgatgacct	ctttatagcc	acctttgttc	atggcagcca	660
gcataatggc	tatgttgcca	aactctaaac	caaatactca	ttctgatgtt	ttaaattgatt	720
tgccctccca	tatgtccttc	cgagttagag	acacaaaaaa	ttccaacaca	ctattgcaat	780
gaaaaataat	ttcctttatt	agccagaagt	cagatgctca	aggggcttca	tgatgtcccc	840
ataatttttg	gcagagggaa	aaagatctca	gtggtatttg	tgagccaggg	cattggccac	900
accagccacc	accttctgat	aggcagcctg	accgtgagga	gtgaattctt	acttgtacag	960
ctcgtccatg	ccgagagtga	tcccggcggc	ggtcacgaac	tccagcagga	ccatgtgatc	1020
gcgctctctg	ttggggctct	tgctcagggc	ggactgggtg	ctcaggtagt	gggtgtcggg	1080
cagcagcagc	ggggcgctgc	cgatgggggt	gttctgctgg	tagtgggtcg	cgagctgcac	1140
gctgcccgtc	tcgatgttgt	ggcggatctt	gaagttcacc	ttgatgccgt	tcttctgctt	1200
gtcggccatg	atatagacgt	tgtggcttgt	gtagttgtac	tccagcttgt	gccccaggat	1260
ggtgcgcttc	tccttgaagt	cgatgccctt	cgctcgatg	cggttcacca	gggtgtcgcc	1320
ctcgaacttc	acctcggcgc	gggtcttgta	gttgccgtcg	tccttgaaga	agatgggtcg	1380
ctcctggacg	tagccttcgg	gcatggcgga	cttgaagaag	tcgtgctgct	tcattgtggtc	1440
ggggtagcgg	ctgaagcact	gcacgcgcta	ggtcagggtg	gtcacagagg	tgggcccagg	1500

-100-

cacggggcagc	ttgccgggtgg	tgcagatgaa	cttcagggtc	agcttgccgt	aggtgggcac	1560
gcccctcgccc	tcgcccggaca	cgctgaactt	gtggccggtt	acgtcgccgt	ccagctcgac	1620
caggatgggc	accaccccgg	tgaacagctc	ctcgcccttg	ctcaccatgg	tggcgaaattc	1680
tttgccaaaa	tgatgagaca	gcacaacaac	cagcacgttg	cccaggagct	gtaggaaaaa	1740
gaagaaggca	tgaacatggg	tagcagaggg	tctagagccg	ccggtcacac	gccagaagcc	1800
gaaccccggc	ctgccccgtc	ccccccgaag	gcagccgtcc	ccctgcggca	gccccgaggc	1860
tggagatgga	gaaggggacg	gcggcgccgg	gacgcacgaa	ggccctcccc	gcccattttcc	1920
ttcctggccg	cgccgcaccc	cttcgcccgc	gcccgcctaga	gggggtgccc	cgccgcctcc	1980
cagatttcgg	ctccgccaga	tttgggacaa	aggaagtccc	tgcccccctc	cgccgcctcc	2040
ccataaaagg	caatggctgc	ggctcgccgc	gcttcgacag	ccgcccggcg	tcggggggccg	2100
ccgcgcccc	cccccgagcc	ctccccggcc	cgaggcgccc	ccgccccggc	cgccaccccc	2160
acctgcccgc	accccccgcc	cgccacggcg	agcccccgcc	cacgccccgc	acggagcccc	2220
gcaccccgaag	ccggggccgtg	ctcagcaact	cgggggagggg	gggtgcagggg	gggggttacag	2280
ccccgacccg	cgcccccacac	cccctgctca	cccccccacg	cacacacccc	gcacaccccc	2340
tttggttccc	tcgcagcccc	ccccgcaccc	ggggcaccgc	ccccggccgc	gctccccctcg	2400
cgcacacggc	gagcgcacaa	agcccccgcc	cgcccccgcg	ggcgtcacag	ccgcccgggca	2460
gcgcggggcg	cacggcgccg	tccccacgca	cacacacacg	cacgaccccc	ccgagcccgct	2520
cccccccgca	caaaggggccc	tccccggagc	ctttaaggct	ttcacgcagc	cacagaaaaag	2580
aaacgagccg	tcaattaaacc	aagcgctaact	tacagcccg	aggagaaagg	ccgtcccggcc	2640
cgctcacctg	tgggagtaac	gcggtcagtc	agagccgggg	cgggcgccgc	gaggcgccgc	2700
ggagcggggg	acggggcgaa	ggcaacgcag	cgactcccgc	ccgcccggcg	cttcgctttt	2760
tataggggcg	ccgcccggcg	cgctcgcca	taaaaggaaa	ctttcgagag	gcgcccgtct	2820
gattggctgc	cgccgcacct	ctccgctcg	ccccccccgc	cccccgccc	cgccccgccc	2880
cgctggcg	gcgccccccc	ccccccccgc	cccatcgctg	cacaaaaata	ttaaaaata	2940
aataaatata	aaattggggg	tggggagggg	ggggagatgg	ggagagtga	gcagaacgtg	3000
gggctccat	cgacccatgg	taatagcgat	gactaatacg	tagatgtact	gccaaagtag	3060
aaagtcacat	aaggtcatgt	actgggcata	atgccaggcg	ggccatttac	cgctattgac	3120
gtcaataggg	ggcgctactg	gcataatgata	cacttgatgt	actgccaaag	gggcagttta	3180
ccgtaaatag	tccacccatt	gacgtcaatg	gaaagtccct	attggcggtta	ctatggggaac	3240
atacgtcatt	attgacgtca	atggggcggg	gctcacaatt	ggtcagccag	cgggggccatt	3300
taccgtaagt	tatgtaacgc	ggaactccat	atatgggcta	tgaactaatg	accccgtaat	3360
tgattactat	taataactag	aggatcccc	ggtagccagc	tcgaattcgt	aatcatgggtc	3420
atagctgttt	cctgtgtgaa	attgttatcc	gtccacaatt	ccacacaaca	tacgagccgg	3480
aagcataaag	tgtaaagcct	ggggtgccta	atgagtgagc	taactcacat	taattgcgtt	3540
gcgctcactg	cccgcctttc	agtcgggaaa	cctgtcgtgc	cagctgcatt	aatgaatcgg	3600
ccaacgcgcg	gggagagggc	gtttgcgtat	tggggcgctc	tcgcgttcc	cgctcactga	3660
ctcgctgcgc	tcggctcggtc	ggctgcggcg	agcgggtatc	gctcactcaa	aggcggtaat	3720
acggttatcc	acagaatcag	gggataacgc	aggaaagaac	atgtgagcaa	aaggccagca	3780
aaagccaggg	aaccgtaaaa	aggccgcgtt	gtcgcgcttt	ttccataggc	tcgccccccc	3840
tgacgagcat	cacaaaaatc	gacgctcaag	tcagaggtgg	cgaaacccga	caggactata	3900
aagataccag	gcgtttcccc	ctggaagctc	cctcgctgcg	tctcctgttc	cgacccctgcc	3960
gcttacccga	tacgtgtccg	cctttctccc	ttcggaagc	gtggcgcttt	ctcatcactg	4020
acgctgtagg	tatctcagtt	cggtgtaggt	cgctcgctcc	aagctgggct	gtgtgcacga	4080
acccccggtt	cagcccgacc	gctgcgcctt	atccggtaac	tatcgtcttg	agtcacaccc	4140
ggtaagacac	gacttatcgc	cactggcagc	agccactggg	aacaggatta	gcagagcgag	4200
gtatgtaggc	gggtgtacag	agttcttgaa	gtgggtggct	aactacggct	acactagaag	4260
gacagtattt	gggtatctgc	ctctgctgaa	gccagttacc	ttcggaaaaa	gagttggtag	4320
ctcttgatcc	ggcaaacaaa	ccaccgctgg	tagcgggtgg	ttttttgttt	gcaagcagca	4380
gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	atcttttcta	cggggtctga	4440
cgctcagtg	aacgaaaaac	cacgttaagg	gatttttggt	atgagattat	caaaaaaggt	4500
cttcacctag	atccttttaa	aagtttttaa	aagtttttaa	tcaatctaaa	gtatatatga	4560
gtaaacttgg	tctgacagtt	accaatgctt	aatcagtgag	gcacctatct	cagcgatctg	4620
tctatttcgt	tcatccatag	ttgcttgact	ccccgtcggt	tagataacta	cgatacggga	4680
gggcttacca	gtctgccaat	gtctgccaat	gaaccgcga	gaccacgct	caccggctcc	4740
agatttatca	gcaataaaac	agccagccgg	aagggccgag	cgcagaagtg	gtcctgcaac	4800
tttatccgce	tccatccagt	ctattaattg	ttgcccggaa	gctagagtaa	gtagttcgcc	4860
agtttaatag	ttgcgcaacg	ttgttgccat	tgctacagcg	atcgtgggtg	cacgctcgct	4920
gtttgggtatg	gcttcattca	gctccgggtc	ccaacgatca	aggcgagtta	catgatcccc	4980
catgtttgtg	aaaaaagcgg	ttagctcctt	cggtcctccg	atcgttgtca	gaagtaagtt	5040
ggccgcagtg	ttaactcaca	tggttatggc	agcactgcac	aattctctta	ctgtcatgcc	5100
atccgtaaga	tgctttttctg	tgactgggtga	gtactcaacc	aagtcattct	gagaatagtg	5160
tatgcggcga	ccgagttgct	cttgcggcgc	gtcaatacgg	gataataccg	cgccacatag	5220
cagaacttta	aaagtgtcca	tatttgga	acgttcttcg	gggcgaaaaa	tctcaaggt	5280
cttaccgctg	ttgagatcca	gttcgatgta	accactcgct	gcacccaact	gatcttcagc	5340
atcttttact	ttcaccagcg	tttctgggtg	agcaaaaaa	ggaaggcaaa	atgccgcaaa	5400
aaagggaaata	agggcgacac	ggaaatgttg	aatactcata	ctcttccttt	ttcaatatta	5460
ttgaagcatt	tatcagggtt	attgtctcat	gagcgatac	atatttgaat	gtatttagaa	5520

-101-

```
aaataaacaa ataggggttc cgcgcacatt tccccgaaaa gtgccacctg acgtagttaa 5580
caaaaaaaag cccgccgaag cgggctttat taccaagcga agcgccattc gccattcagg 5640
ctgcgcaact gttgggaagg gcgatcgggt cgggcctcct cgctattacg ccagctggcg 5700
aaagggggat gtgctgcaag gcgattaagt tgggtaacgc caggggtttc ccagtcacga 5760
cgttgtaaaa cgacggccag tccgtaatac gactcactta aggccttgac tagaggggtcg 5820
acggtataca gacatgataa gatacattga tgagtttgga caaacacaa ctagaatgca 5880
gtgaaaaaaa tgcttttatt gtgaaatttg tgatgctatt gctttatttg taaccattat 5940
aagctgcaat aaacaagttg gggtagggcga agaactccag catgagatcc ccgcgctgga 6000
ggatcatcca gccggcgtcc cggaaaacga ttccgaagcc caacctttca tagaaggcgg 6060
cggtggaatc gaaatctcgt agcagtgctc agtctctgctc ctcgccacag aagtgcacg 6119
```

&lt;210&gt; 127

&lt;211&gt; 5855

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; pCXLamInt Plasmid (Wildtype Integrase)

&lt;400&gt; 127

```
gtcgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60
gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgccct ggctgaccgc 120
ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180
ggactttcca ttgacgtcaa tgggtggact atttacggta aactgcccac ttggcagtac 240
atcaagtgtg tcatatgcca agtacgcccc ctattgacgt caatgacggg aaatggcccc 300
cctggcatta tgcccagtac atgaccttat gggactttcc tacttggcag tacatctacg 360
tattagtcac cgctattacc atgggtcgag gtgagcccca cgttctgctt cactctcccc 420
atctcccccc cctccccacc cccaattttg tatttattta ttttttaatt attttgtgca 480
gcgatggggg cggggggggg gggggcgcg cccaggcggg gcggggcggg gcgagggggc 540
gggcggggcg aggcggagag gtgcggcggc agccaatcag agcgcgcgcc tccgaaagt 600
tccttttatg gcgaggcggc ggcggcggcg gccctataaa aagcgaagcg cgcggcgggc 660
gggagtcgct gcgttgccct cgcgccgtgc cccgctccgc gccgcctcgc gccgcccgcg 720
ccggtctga ctgaccgcgt tactcccaca ggtgagcggg cgggacggcc cttctcctcc 780
gggctgtaat tagcgcttgg tttaatgacg gctcgtttct tttctgtggc tgcgtgaaag 840
ccttaaaggg ctccgggagg gcccttttgt cgggggggag cggctcgggg ggtgctgctg 900
tgtgtgtgtg cgtggggagg gcccgctgcg ccccgcgctg cccggcggtg gtgagcgctg 960
cgggcgcggc gcggggcctt gtgcgctccg cgtgtgcgcg aggggagcgc ggccgggggc 1020
ggtgccccgc ggtgcggggg ggctgcgagg ggaacaaagg ctgcgtgcgg ggtgtgtgcg 1080
tgaggggggt agcagggggg gtgggcgcgg cgttcgggct gtaaccccc cctgcaccgc 1140
cctccccgag ttgctgagca cggcccggtc tcgggtgcgg ggtccgtgc gggtcgtggc 1200
gcggggctcg ccgtgcgggg cgggggggtg cggcaggtgg ggggtgcggg cggggcgggg 1260
ccgctctggg ctcggggagg ctcggggggg gggcgcgccg gccccggagc gcccgcgct 1320
gtcgaggcgc ggogagcgcc agccattgcc ttttatggta atcgtgcgag agggcgccag 1380
gacttccttt gtcccaaacc tggcggagcc gaaatctggg aggcgcgcgc gcacccccct 1440
tagcgggcgc gggcgaaagc gtgcggcgcc ggcaggaagg aaatggggcg ggagggcctt 1500
cgtgcgtcgc cgcgcgcgcg tccccttctc catctccagc ctcggggctg ccgcaggggg 1560
acggctgcct tcggggggga cggggcaggg cgggggtcgg cttctggcgt gtgaccggcg 1620
gctctagagc ctctgctaac catgttcatt cttcttctt tttcctacag ctctggggca 1680
acgtgctggt tgttgtgtgt tctcatcatt ttggcaaaga attcatggga agaaggcgaa 1740
gtcatgagcg ccgggattta cccoctaacc tttatataag aaacaatgga tattactgct 1800
acagggaccc aaggacgggt aaagagtttg gattaggcag agacaggcga atcgcaatca 1860
ctgaagctat acaggccaac attgagttat tttcaggaca caaacacaag cctctgacag 1920
cgagaatcaa cagtgataat tocgttacgt tacattcatg gcttgatcgc tacgaaaaaa 1980
tcctggccag cagaggaatc aagcagaaga cactcataaa ttacatgagc aaaattaaag 2040
caataaggag gggctctgct gatgctccac ttgaagacat caccacaaaa gaaattgcgg 2100
caatgctcaa tggatacata gacgagggca aggcggcgct agccaagtta atcagatcaa 2160
cactgagcga tgcattccga gaggcaatag ctgaaggcca tataacaaca aacctgtcgc 2220
ctgccactcg cgcagcaaaa tcagaggtaa ggagatcaag acttacggct gacgaatacc 2280
tgaaaaattt tcaagcagca gaatcatcac catgttggct cagacttgca atggaactgg 2340
ctgttgttac cgggcaacga gttgggtgatt tatgcgaaat gaagtggctc gatatcgtag 2400
atggatatct ttatgtcgag caaagcaaaa caggcgtaaa aattgccatc ccaacagcat 2460
tgcatattga tgctctcgga atatcaatga aggaacact tgataaatgc aaagagattc 2520
ttggcgggga aaccataatt gcatctactc gcgcgaacc gctttcatcc ggcacagtat 2580
caaggtatct tatgcgcgca cgaaaagcat caggctcttc cttcgaaggc gatccgcta 2640
cctttcacga gttgcgcagt ttgtctgcaa gactctatga gaagcagata agcgataagt 2700
ttgtctcaac tcttctcggg cataagtcgg acaccatggc atcacagtat cgtgatgaca 2760
gaggcagggg gtgggacaaa attgaaatca aataagaatt cactcctcag gtgcaggctg 2820
```

-102-

```

cctatcagaa ggtggtggct ggtgtggcca atgccctggc tcacaaatac cactgagatc 2880
ttttccctc tgccaaaaat tatggggaca tcatgaagcc ccttgagcat ctgacttctg 2940
gctaataaag gaaatatttatt ttcatgtcaa tagtgtgttg gaatttttttg tgtctctcac 3000
tcggaaggac atatgggagg gcaaatcatt taaaacatca gaatgagtat ttgggttaga 3060
gtttggcaac atatatgccata tgctggctgc catgaacaaa ggtggctata aagaggtcat 3120
cagtatatga aacagccccc tgctgtccat tccttattcc atagaaaagc cttgacttga 3180
ggttagattt tttttatatt ttgttttgtg ttattttttt ctttaacatc cctaaaattt 3240
tccttacatg ttttactagc cagattttttc ctctctcctt gactactccc agtcatagct 3300
gtccctcttc tcttatgaag atccctcgac ctgcagccca agcttggcgt aatcatggtc 3360
atagctgttt cctgtgtgaa attgttatcc gctcacaatt ccacacaaca tacgagccgg 3420
aagcataaag tgtaaaagcct ggggtgccta atgagtggc taactcacat taattgctgt 3480
gcgctcactg ccgcctttcc agtcgggaaa cctgtcgtgc cagcggatcc gcactctaat 3540
tagtcagcaa ccatagtccc gcccttaact ccgcccattcc cgccttaac tccgccagt 3600
tccgccattt ctccgcccc ctgagctatt tggctgacta attttttta tttatgcaga ggccgaggcc 3660
gctcggcctt ctgagctatt taacttgttt ccagaagtag tgaggaggct tttttggagg cctaggcttt 3720
tgcaaaaagc caaatttcac aaataaagca tttttttcac tgaggaggct caaaataaagc aatagcatca 3780
tcaatgtatc ttatcatgtc tggtatccgt gcattaatga atcggccaac atcggccaac ggcgggggag 3900
aggcggtttg cgtattgggc gctcttccgc ttccctcgct ctcctcgctc actgactcgc tgcgtcgggt 3960
cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggg taatccacaga 4020
atcaggggat aacgcaggaa agaacatgtg agcaaaaagg cagcaaaaagg ccaggaaccg 4080
taaaaaggcc gcgttgctgg cgtttttcca taggctccgc ccccttgacg agcatcacia 4140
aaatcgacgc tcaagtcaga ggtggcgaaa cccgcagga ctataaagat accaggcgtt 4200
tccccctgga agctccctcg tgctctctcc aatttccgac ctgcccgtta cgggatacct 4260
gtccgccttt ctcccttccg gaagcgtggc gcttccagct gtttccgacc tgctcacgct gttagtatct 4320
cagttcggtg taggtcgttc gtaactatcg ctggttaacag gggctgtgtg cactgaacccc cgttcagcc 4380
cgaccgtgct gcagcagcca ttgaagtggg ttaccttcgg ggtagctctt gatccggcaa 4440
atcgccactg gcagcagcca gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgcgagaaa 4500
tacagagttc ctgaagccag gctggtagcg gtggtttttt ttctacgggg tctgacgctc agtggaaaca 4560
ctgctctctg ctgaagccag gctggtagcg gtggtttttt attatcaaaa aggatcttca cctagatcct 4620
acaaaccacc gctggtagcg caagaagatc taagggattt ttaaatacaat cttaaagtata tatgagtaaa 4680
aaaaggatct aaactcagct aaatgaagtt tgcttaatca atctgtctat atctgtctat ttcgttcac 4740
tttaaattaa tgcttaatca tgctgtatag tctgtatagat cgggaggggt taccatctgg 4800
cagttaccac catagttgcc gcaatgatac ccgagcgcag gggagagctag cgtccagatt tatcagcaat 4860
ccagctcatt aattgttgcc gccattgcta ggttcccaac tcttcgggtc atagtttgag 4920
caacgttgtt attcagctcc agcggttagc atggcagcac ttctgtgact ggtgagtagt ccggcgtaaa 4980
agcggttagc tcttcgggtc atggcagcac ttctgtgact ggtgagtagt ccggcgtaaa 5040
actcatgggt ttctgtgact ggtgagtagt ccggcgtaaa 5100
ttctgtgact ggtgagtagt ccggcgtaaa 5160
ttctgtgact ggtgagtagt ccggcgtaaa 5220
ttctgtgact ggtgagtagt ccggcgtaaa 5280
ttctgtgact ggtgagtagt ccggcgtaaa 5340
ttctgtgact ggtgagtagt ccggcgtaaa 5400
ttctgtgact ggtgagtagt ccggcgtaaa 5460
ttctgtgact ggtgagtagt ccggcgtaaa 5520
ttctgtgact ggtgagtagt ccggcgtaaa 5580
ttctgtgact ggtgagtagt ccggcgtaaa 5640
ttctgtgact ggtgagtagt ccggcgtaaa 5700
ttctgtgact ggtgagtagt ccggcgtaaa 5760
ttctgtgact ggtgagtagt ccggcgtaaa 5820
ttctgtgact ggtgagtagt ccggcgtaaa 5880

```

&lt;210&gt; 128

&lt;211&gt; 303

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Human FER-1 Promoter

&lt;400&gt; 128

```

tccatgacaa agcacttttt gagcccaagc ccagcctagc tcgagctaaa cgggacagaa 60
gacgcacacg ctgtcccaga ggcagtcggc taccggtccc cgctcccag ccccgccaga 120
gcgcgcagcg gccctccagcg gccgcccctc ccccacagca gggcgggggg ccccgcccca 180
ccggaaggag cgggctcggg gcggcgggcg ctgattggcc gggcgggggc tgacgcggac 240
gcggctataa gagaccacaa gcgaccgcga gggccagacg ttcttcgccc agagtcgggt 300
acc

```

-103-

<210> 129  
<211> 6521  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> pIRES-BSR Plasmid

<400> 129  
tcaatattgg ccattagcca tattattcat tgggtatata gcataaatca atattggcta 60  
ttggccattg catacgttgt atctatatca taatatgtac atttatattg gctcatgtcc 120  
aatatgaccg ccatgttggc attgattatt gactagttat taatagtaat caattacggg 180  
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc 240  
gctggctga ccgcccacg acccccgccc attgacgtca ataatgacgt atgttcccat 300  
agtaacgcca atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaactgc 360  
ccacttggca gtacatcaag tgtatcata gccaaagtcg cccctattg acgtcaatga 420  
cggtaaatgg ccgcctggc attatgccc gtacatgacc ttacgggact ttcctacttg 480  
gcagtacatc tacgtattag tcatcgctat taccatgggt atgcgggttt ggcagtacac 540  
caatgggctg ggatagcgt ttgactcacg gggatttcca agtctccacc ccattgacgt 600  
caatgggagt ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaacaactg 660  
cgatcgcccg cccgttgac gcaaatgggc ggtaggcgtg tacgggtggga ggtctatata 720  
agcagagctc gtttagtgaa ccgtcagatc actagaagct ttattgcggt agtttatcac 780  
agttaaattg ctaacgcagt cagtgttctt gcagaaagtg tctcgaactt aagctgcagt 840  
gactctctta aggtagcctt gcagaagtgg gctgtgaggg actgggcagg taagtatcaa 900  
ggttacaaga caggtttaag gagaccaata gaaactgggc ttgtcgagac agagaagact 960  
cttgcgtttc tgataggcac ctattggctc tactgacatc cactttgcct ttctctccac 1020  
aggtgtccac tcccagttca attacagctc ttaaggctag agtacttaac acgactcact 1080  
ataggctagc ctcgagaatt caccgctcga gcatgcatct agggcgccca attccgcccc 1140  
tctccctccc cccccctaa cgttactggc cgaagccgct tgggaataagg ccggtgtgcg 1200  
tttgtctata tgtgattttc caccatattg ccgtcttttg gcaatgtgag ggcccggaaa 1260  
cctggccctg tcttcttgac gagcattcct aggggtcttt cccctctcgc caaaggaatg 1320  
caaggtctgt tgaatgtcgt gaaggaagca gttcctctgg aagcttcttg aagacaaaac 1380  
acgtctgtag cgaccctttg caggcagcgg aaccccccac ctggcgacag gtgcctgc 1440  
ggccaaaagc cacgtgtata agatacacct gcaaaggcgg cacaacccca gtgccacgtt 1500  
gtgagttgga tagttgtgga aagagtcaaa tggctctcct caagcgtatt caacaagggg 1560  
ctgaaggatg acaagattct accccattgt atgggtctg atctggggcc tggctcctgc 1620  
tgctttacat gtgttttagt gaggttaaaa aaacgtctag gccccccgaa ccacggggac 1680  
gtggttttcc tttgaaaaac acgatgataa gcttgccaca acccaccatg aaaacattta 1740  
acatttctca acaagattct gaattagtag aagtagcgac agagaagatt acaatgcttt 1800  
atgaggataa taaacatcat gtgggagcgg caattcgtac gaaaacagga gaaatcattt 1860  
cggcagtaaa tattgaagcg tatataggac gagtaactgt ttgtgcagaa gccattgcga 1920  
ttgttagtgc agtctcgaat ggacaaaagg attttgacac gattgtagct gtttagcacc 1980  
cttattctga cgaagtagat agaagtattc gagtggttaag tccttgtggg atgtgtaggg 2040  
agttgatttc agactatgca ccagattggt ttgtgttaat agaaatgaat ggcaagttag 2100  
tcaaaactac gattgaagaa ctcatctccac tcaaatatac ccgaaattaa aagttttacc 2160  
ataccaagct tggcggggcg cctcttccct ttagtgaggg ttaatgcttc gagcagacat 2220  
gataagatac attgatgagt ttggacaaac cacaactaga atgcagtga aaaaatgctt 2280  
tatttgtgaa atttgtgatg atttghtaacc attataagct gcaataaaca 2340  
agttaacaac aacaattgca ttcattttat gtttcagggt caggggggaga tgtggggagg 2400  
tttttaaaag aagtaaaacc tctacaaatg tggtaaaatc cgataaggat cgatccgggc 2460  
tggcgtaata gcgaagaggc cgcaccgat cgcccttccc aacagttgcg cagcctgaat 2520  
ggcgaatgga cgcgccctgt agcggcgcgt taagcgcggc ggggtgtggg gttacgcgca 2580  
gcgtgaccgc tacacttgcc agcgccttag cgcccgctcc ttctcgcttc tcccttccct 2640  
ttctcgccac gttcgccggc tttcccgtc aagctctaaa tcgggggctc cttttagggg 2700  
tccgatttag agctttacgg cacctcgacc gcaaaaaact tgatttgggt gatggttcac 2760  
gtagtggggc atcgccctga tagacgggtt ttccgctttt gacgttggag tccacgttct 2820  
ttaatagtgg actcttggtc caaactggaa caacactcaa cctattctcg gtctattctt 2880  
ttgatttata agggattttg ccgatttcgg cctattgggt aaaaaatgag ctgatttaac 2940  
aaatatttaa cgcgaatttt aacaaaatct taacggtttac aatttcgcct gatgcggtat 3000  
tttctcctta gcactctgtg cggatttcca caccgcatac gcggtctgca gcaagaacc 3060  
ggcctgaaat aacctctgaa agaggaactt ggttaggtac cttctgaggg ggaaagaacc 3120  
agctgtggaa tgtgtgtcag ttaggggtgt gaaagctccc aggtctccca gcaggcagaa 3180  
gtatgcaaaag catgcatctc aattagtcag caaccaggtg tggaaaagtcc ccaggctccc 3240  
cagcaggcag aagatgcaa agcatgcat tcaattagtc agcaaccata gtcccgcccc 3300  
taactccgcc catccgccc ctaactccgc ccagttccgc ccattctccg ccccatgggt 3360  
gactaatatt ttttatttat gcagaggcgg aggcgcctc ggccctctgag ctattccaga 3420  
agtagtgagg aggccttttt gtagggcctag gcttttgcaa aaagcttgat tcttctgaca 3480

-104-

caacagtcctc	gaacttaagg	ctagagccac	catgattgaa	caagatggat	tgcacgcagg	3540
ttctccggcc	gcttgggtgg	agaggctatt	cggtatgac	tgggcacaa	agacaatcgg	3600
ctgctctgat	gcgcgcgtgt	tccggctgtc	agcgcagggg	cgcccggttc	tttttgtcaa	3660
gaccgacctg	tccggtgccc	tgaatgaact	gcaggacgag	gcagcgcggc	tatcgtggct	3720
ggccacgacg	ggcgttcctt	gcgcagctgt	gctcgacgtt	gtcactgaag	cggaagggga	3780
ctggctgcta	ttgggcgaag	tgcgggggca	ggatctcctg	tcatctcacc	ttgctcctgc	3840
cgagaaagta	tccatcatgg	ctgatgcaat	gcggcggtcg	catacgcttg	atccggctac	3900
ctgccccattc	gaccaccaag	cgaaacatcg	catcgagcga	gcacgtactc	ggatggaagc	3960
cggtcttgtc	gatcaggatg	atctggacga	agagcatcag	gggctcgcg	cagccgaact	4020
gttcgccagg	ctcaaggcgc	gcattgccga	cggcgaggat	ctcgtcgtga	cccatggcga	4080
tgcctgcttg	ccgaatatca	tgggtgaaaa	tggccgcttt	tctggattca	tcgactgtgg	4140
ccggtgggt	gtggcgacc	gctatcagga	catagcgttg	gctaccgctg	atattgctga	4200
agagcttggc	ggcgaatggg	ctgaccgctt	cctcgtgctt	tacggtatcg	ccgctcccga	4260
ttcgcagcgc	atcgcttctt	atcgcttctt	tgacgagttc	ttctgagcgg	gactctgggg	4320
ttcgaaatga	ccgaccaagc	gacgcccac	ctgccatcac	gatggccgca	ataaaatata	4380
tttatttttca	ttacatctgt	gtgttggttt	tttgtgtgaa	tcgatagcga	taaggatccg	4440
cgtatgggtgc	actctcagta	caatctgctc	tgatgccgca	tagttaagcc	agccccgaca	4500
cccgccaaaca	cccgtgacg	cgccctgacg	ggcttgtctg	ctcccggcat	ccgcttacag	4560
acaagctgtg	accgtctccg	ggagctgcat	gtgtcagagg	ttttcacctg	catcaccgaa	4620
acgcgcgaga	cgaaagggcc	tcgtgatagc	cctattttta	taggttaatg	tcattgataat	4680
aatggtttct	tagacgtcag	gtggcacttt	tcggggaaat	gtgcgcggaa	cccctatttg	4740
tttattttttc	taaatacatt	caaatatgta	tccgctcatg	agacaataac	cctgataaat	4800
gcttcaataaa	tattgaaaaa	ggaagagtat	gagtatccaa	catttccgtg	tcgccccttat	4860
tcccttttttt	gcggcatttt	gccttctctg	ttttgctcac	ccagaaacgc	tggtgaaagt	4920
aaaagatgct	gaagatcagt	tgggtgcacg	agtgggttac	atcgaactgg	atctcaacag	4980
cggtaaagatc	cttgagagtt	ttcgccccga	agaacgtttt	ccaatgatga	gcacttttaa	5040
agttctgcta	tgtggcgcg	tattatcccg	tattgacgcc	gggcaagagc	aactcggtcg	5100
ccgcatacac	tattctcaga	atgacttggg	tgagtactca	ccagtcacag	aaaagcatct	5160
tacggatggc	atgacagtaa	gagaattatg	cagtgtgccc	ataaccatga	gtgataacac	5220
tgcggccaac	ttactttctga	caacgatcgg	aggaccgaag	gagctaaccg	cttttttgca	5280
caacatgggg	gatcatgtaa	ctcgcttga	tcgttgggaa	ccggagctga	atgaagccat	5340
accaaaccgac	gagcgtgaca	ccacgatgcc	tgtagcaatg	gcaacaacgt	tgcgcaaaact	5400
attaactggc	gaactactta	ctctagcttc	ccggcaacaa	ttaatagact	ggatggaggc	5460
ggataaaagtt	gcaggaccac	ttctgcgctc	ggcccttccg	gctggctggg	ttattgctga	5520
taaatctgga	gccggtgagc	gtgggtctcg	cggtatcatt	gcagcactgg	ggccagatgg	5580
taagccctcc	cgtatcgtag	ttatctacac	gacggggagt	caggcaacta	tggatgaacg	5640
aaatagacag	atcgctgaga	taggtgcctc	actgattaag	cattggtaac	tgtcagacca	5700
agtttactca	tatatacttt	agattgattt	aaaacttcat	ttttaattta	aaaggatcta	5760
ggtgaagatc	cttttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcgttcca	5820
ctgagcgtca	gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	tttttctgcg	5880
cgtaatctgc	tgcttgcaaa	caaaaaaacc	accgctacca	gcgggtgggtt	gtttgccgga	5940
tcaagagcta	ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	6000
tactgtcctt	ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcaccgcc	6060
tacatacctc	gctctgctaa	tcctgttacc	agtggctgct	gccagtggcg	ataagtctgt	6120
tcttaccggg	tgggactcaa	gacgatagtt	accggataag	gcgcagcggg	cgggctgaac	6180
gggggggttcg	tgcacacagc	ccagcttgga	gcgaacgacc	tacaccgaac	tgagatacct	6240
acagcgtgag	ctatgagaaa	gcgccacgct	tcccgaaggg	agaaaaggcg	acaggtatcc	6300
ggtaagcggc	agggctcgaa	caggagagcg	cacgagggag	cttccagggg	gaaacgcctg	6360
gtatctttat	agtcctgtcg	ggtttcgcca	cctctgactt	gagcgtcgat	ttttgtgatg	6420
ctcgtcaggg	gggcggagcc	tatggaaaaa	cgccagcaac	gcggcctttt	tacggttcct	6480
ggccttttgc	tggccttttg	ctcacatggc	tcgacagatc	t		6521

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number  
WO 02/097059 A3

(51) International Patent Classification<sup>7</sup>: C07H 21/04,  
C12N 15/85, 15/87, 15/90, A01N 43/34, C12N 15/09

(21) International Application Number: PCT/US02/17452

(22) International Filing Date: 30 May 2002 (30.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/294,758 30 May 2001 (30.05.2001) US  
60/366,891 21 March 2002 (21.03.2002) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:  
US 60/294,758 (CIP)  
Filed on 30 May 2001 (30.05.2001)  
US 60/366,891 (CIP)  
Filed on 21 March 2002 (21.03.2002)

(71) Applicant (for all designated States except US): CHROMOS MOLECULAR SYSTEMS, INC. [CA/CA]; 8081 Loughheed Highway, Burnaby, B.C. V5A 1W9 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): PERKINS, Edward [US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6 (CA). PEREZ, Carl [US/CA]; 1201-7680 Granville Avenue, Richmond, B.C. V6Y 4B9 (CA). LINDENBAUM, Michael [CA/CA]; 252 Finnigan Street, Coquitlam, B.C. V3K 5J7 (CA). GREENE, Amy [US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6 (CA). LEUNG, Josephine [CA/CA]; 711 Ebert Avenue, Coquitlam, B.C. V3J 7P8 (CA). FLEMING, Elena [CA/CA]; 248 E 18th, North Vancouver, B.C. V7L 2X6 (CA). STEWART, Sandra [CA/CA]; 2618 Oxford Street, Vancouver, B.C. V5K 1N3 (CA). SHELLARD, Joan [CA/CA]; #215-1345 West 15th Avenue, Vancouver, B.C. V6H 3R3 (CA).

(74) Agents: SEIDMAN, Stephanie, L. et al.; Heller Ehrman White & McAuliffe LLP, 7th floor, 4350 La Jolla Village Drive, San Diego, CA 92122-1246 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,

MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

— with international search report

(88) Date of publication of the international search report:  
30 May 2003

[Continued on next page]

(54) Title: CHROMOSOME-BASED PLATFORMS

(57) Abstract: Artificial chromosomes, including *Aces*, that have been engineered to contain available sites for site-specific, recombination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational engineering of the chromosome for a variety of applications.

WO 02/097059 A3

**WO 02/097059 A3**

---



*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/17452

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/04; C12N 15/85,87,90; A01N 43/34; C12N 15/09

US CL : 536/23.1; 435/320.1,325,419,455,467; 514/44; 800/21

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1; 435/320.1,325,419,455,467; 514/44; 800/21

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
APS (EAST); STN (MEDLINE, BIOSIS, CAPLUS)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 01/07572 A2 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 01 February 2001 (01.02.2001), see entire document.	1,2,7,8,11,23- 25,27,50 ----- 3-6,9,10,12,13,26,28- 1,2,7,8,11,23- 25,27,50
X --- Y	WO 00/11155 A1 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 02 March 2000 (02.03.2000), see entire document.	----- 3-6,9,10,12,13,26,28- 1-3,8,10-12,23- 27,33,36,39,40,43- 45,50,54,55,85 ----- 4-7,9,13-22,28- 30,32,34,35,37,38,41, 42,46-49,51-53,56,67- 71,84,86
X --- Y	US 6,171,861 B1 (HARTLEY et al.) 09 January 2001 (09.01.2001), see entire document, especially sequence listing.	

☒ Further documents are listed in the continuation of Box C.☐ See patent family annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"B" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&amp;"

document member of the same patent family

Date of the actual completion of the international search

01 November 2002 (01.11.2002)

Date of mailing of the international search report

27 NOV 2002

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Daniel M Sullivan

Telephone No. 703-308-0196

# INTERNATIONAL SEARCH REPORT

PCT/US02/17452

## C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 96/40724 A1 (LIFE TECHNOLOGIES, INC.) 19 December 1996 (19.12.96), see entire document, especially SEQ ID NO:1-16.	1-3,8,10-12,23- 27,33,36,39,40,43- 45,50,54,55,85 ----- 1-3,8,10-12,23- 27,33,36,39,40,43- 45,50,54,55,85
Y	WO 94/00569 A1 (GENPHARM INTERNATIONAL, INC.) 06 January 1994 (06.01.94), see entire document.	72-78
Y	WO 94/23049 A2 (THE JOHNS HOPKINS UNIVERSITY) 13 October 1994 (13.10.94), see entire document.	72-78
X --- Y	US 5,721,118 A (SCHEFFLER) 24 February 1998 (24.02.98), see entire document, especially Example 3.	14,15,22-25,57 ----- 16-21,26,58-64,72- 78,84,85,98-105,123
X --- Y	US 5,948,653 A (PATI et al.) 07 September 1999 (07.09.99), see entire document, especially the first full paragraph in column 29.	14,15,22-25,57 ----- 16-21,26,58-64,72- 78,84,85,98-105,123
X --- Y	GB 2 331 752 A (MEDICAL RESEARCH COUNCIL) 02 June 1999 (02.06.99), see entire document, especially pages 8-10.	109,123 ----- 72-78,110-122
A	HALDIMANN et al. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. J Bacteriol. November 2001, Vol. 183 No. 21, pages 6384-6393, see entire document.	65,66
X,P --- Y,P	MORALLI et al. Insertion of a loxP site in a size-reduced human accessory chromosome. Cytogenet. Cell Genet. 2001, Vol. 94 No. 3-4, pages 113-120, see entire document.	91-93,106,108 ----- 94-97,107
X --- Y	LORBACH et al. Site-specific recombination in human cells catalyzed by phage lambda integrase mutants. J. Mol. Biol. March 2000, Vol. 296 No. 5, pages 1175-1181, see entire document.	1,2,8,50 ----- 3-7,9-13,23-32,51- 56,84,85
X --- Y	CALL et al. A cre-lox recombination system for the targeted integration of circular yeast artificial chromosomes into embryonic stem cells. Hum. Mol. Genet. July 2000, Vol. 9, No. 12, pages 1745-1751, see entire document.	14,15,22-26,57,59 ----- 16- 21,58,60,84,85,98- 105
Y,P	HADLACZKY, G. Satellite DNA-based artificial chromosomes for use in gene therapy. Curr. Opin. Mol. Ther. April 2001, Vol. 3 No. 2, pages 125-132, see entire document.	1-123

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/17452

### Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

### Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:  
Please See Continuation Sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.

## INTERNATIONAL SEARCH REPORT

PCT/US02/17452

### BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-64, 67-71, 79, 84-86, 91-108, and 123, drawn to eukaryotic recombinogenic chromosomes used for introducing heterologous nucleic acids into a chromosome and the resulting cells.  
Group II, claim(s) 65-66 and 87-89, drawn to a lambda intR mutain.  
Group III, claim(s) 72-78, drawn to the production of transgenic animals.  
Group IV, claim(s) 80, drawn to the production of an artificial chromosome library.  
Group V, claim(s) 81-83, drawn to a library of cells for genomic screening.  
Group VI, claim(s) 89-90, drawn to a modified iron-induced promoter.  
Group VII, claim(s) 109-122, drawn to a method for screening compounds and their effects on regulatory regions.

and it considers that the International Application does not comply with the requirements of unity of invention (Rules 13.1, 13.2 and 13.3) for the reasons indicated below:

The inventions listed as Groups I-VII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical feature of Group I which defines an advance over the art is a eukaryotic chromosome containing recombinogenic sites that can be used to introduce heterologous nucleic acids into chromosomes, and cells containing these recombinogenic chromosomes.

The special technical feature of Group II involves a lambda intR mutain. This feature defines an advance over Group I in that it involves a protein that is not required for the technical features as set forth above in Group I.

The special technical feature of Group III involves the production of a transgenic animal, which represents a second method for using the invention as set forth above in Group I.

The special technical feature of Group IV involves the production of artificial chromosome expression system libraries, which represents a third method for using the invention as set forth above in Group I.

The special technical feature of Group V involves a library of cells containing the artificial chromosome expression system libraries set forth above in Group IV, and represents a first product resulting from Group IV.

The special technical feature of Group VI involves a modified iron-inducible promoter. This feature defines an advance over Group I in that it involves a promoter that is not required for the technical features as set forth above in Group I.

The special technical feature of Group VII involves a method for screening compounds for their effect on regulatory regions, which represents a fourth method of using the invention as set forth above in Group I.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number  
**WO 2002/097059 A3**

(51) International Patent Classification<sup>7</sup>: **C07H 21/04**,  
C12N 15/85, 15/87, 15/90, A01N 43/34, C12N 15/09

(21) International Application Number:  
PCT/US2002/017452

(22) International Filing Date: 30 May 2002 (30.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/294,758 30 May 2001 (30.05.2001) US  
60/366,891 21 March 2002 (21.03.2002) US

(63) Related by continuation (CON) or continuation-in-part  
(CIP) to earlier applications:  
US 60/294,758 (CIP)  
Filed on 30 May 2001 (30.05.2001)  
US 60/366,891 (CIP)  
Filed on 21 March 2002 (21.03.2002)

(71) Applicant (for all designated States except US): **CHROMOS MOLECULAR SYSTEMS, INC.** [CA/CA]; 8081  
Lougheed Highway, Burnaby, B.C. V5A 1W9 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **PERKINS, Edward**  
[US/US]; 4203 Robinson Street, Duluth, MN 55804 (US). **PEREZ, Carl** [US/CA]; 95 West 11th Avenue, Vancouver,  
British Columbia V5Y 1S6 (CA). **LINDENBAUM, Michael** [CA/CA]; 9941 Martin Court, Burnaby, British  
Columbia V3K 1J5 (CA). **GREENE, Amy** [US/US];  
4203 Robinson Street, Duluth, MN 55804 (US). **LEUNG, Josephine** [CA/CA]; 711 Ebert Avenue, Coquitlam, B.C.  
V3J 7P8 (CA). **FLEMING, Elena** [CA/CA]; 427 Mon-  
troyal Blvd., North Vancouver, British Columbia V7N  
3E2 (CA). **STEWART, Sandra** [CA/CA]; 2618 Oxford  
Street, Vancouver, B.C. V5K 1N3 (CA). **SHELLARD, Joan** [CA/CA]; #215-1345 West 15th Avenue, Vancouver,  
B.C. V6H 3R3 (CA).

(74) Agents: **SEIDMAN, Stephanie, L.** et al.; Heller Ehrman  
White & McAuliffe LLP, 7th floor, 4350 La Jolla Village  
Drive, San Diego, CA 92122-1246 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,

LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,  
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,  
VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,  
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent  
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,  
NE, SN, TD, TG).

**Declarations under Rule 4.17:**

— as to the applicant's entitlement to claim the priority of the  
earlier application (Rule 4.17(iii)) for the following desig-  
nations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY,  
BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC,  
EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN,  
IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV,  
MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH,  
PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR,  
TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent  
(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB,  
GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ,  
CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,  
TG)

— as to the applicant's entitlement to claim the priority of the  
earlier application (Rule 4.17(iii)) for the following desig-  
nations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY,  
BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC,  
EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN,  
IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV,  
MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH,  
PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR,  
TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent  
(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB,  
GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ,  
CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,  
TG)

**Published:**

— with international search report  
— with amended claims

(88) Date of publication of the international search report:  
30 May 2003

[Continued on next page]

(54) Title: **CHROMOSOME-BASED PLATFORMS**

(57) Abstract: Artificial chromosomes, including *Aces*, that have been engineered to contain available sites for site-specific, re-  
combination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational  
engineering of the chromosome for a variety of applications.



WO 2002/097059 A3



**Date of publication of the amended claims:** 31 December 2003

**(15) Information about Correction:**

**Previous Correction:**

see PCT Gazette No. 39/2003 of 25 September 2003, Section II

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

**AMENDED CLAIMS**

[received by the International Bureau on 27 January 2003 (27.01.2003);  
Original claims 22, 29, 39, 40, 68, 104, 107, 108, 114 and 121 replaced by  
Amended Claims 22, 29, 39, 40, 68, 104, 107, 108, 114 and 121.;  
Remaining claims unchanged]

13. The chromosome of claim 6 that is an artificial chromosome expression system (*ACes*).
14. A platform artificial chromosome expression system (*ACes*) comprising one or a plurality of sites that participate in recombinase catalyzed recombination.
15. The *ACes* of claim 14 that contains one site.
16. The *ACes* of claim 14 that is predominantly heterochromatin.
17. The *ACes* of claim 14 that contains no more than about 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.
18. The *ACes* of claim 14 that is a plant *ACes*.
19. The *ACes* of claim 14 that is an animal *ACes*.
20. The *ACes* of claim 14 that is selected from a fish, insect, reptile, amphibian, arachnid or a mammalian *ACes*.
21. The *ACes* of claim 14 that is a fish *ACes*.
22. The artificial chromosome expression system (*ACes*) of claim 14, wherein the recombinase and site(s) are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase system of phage Mu, the Cin recombinase system, the Pin recombinase system of *E. coli*, the R/RS system of the pSR1 plasmid, or any combination thereof.
23. A method of introducing heterologous nucleic acid into a chromosome, comprising:  
contacting a chromosome of any of claims 1 or 14 with a nucleic acid molecule comprising both the heterologous nucleic acid and a recombination site, in the presence of a recombinase that promotes recombination between the sites in the chromosome and in the nucleic acid molecule.

24. The method of claim 23, wherein the recombinase is selected from the group consisting of Cre, Gin, Cin, Pin, FLP, a phage integrase and R from the pSR1 plasmid.

25. The method of claim 23, wherein the nucleic acid molecule  
5 encodes a therapeutic protein, antisense nucleic acid, or comprises an artificial chromosome.

26. The method of claim 25, wherein the nucleic acid molecule comprises a yeast artificial chromosomes (YAC), a bacterial artificial chromosome (BAC) or an insect artificial chromosome (IAC).

10 27. A combination, comprising, the chromosome of claim 1 and a first vector comprising the cognate recombination site, wherein the cognate recombination site is a site that recombines with the site engineered into the chromosome.

15 28. The combination of claim 27, further comprising nucleic acid encoding a recombinase, wherein the nucleic acid is on a second vector or on the first vector, or on the ACes under an inducible promoter.

29. The combination of claim 28, wherein the recombinase and sites are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase  
20 system of phage Mu, the Pin recombinase system of *E. coli*, the R/RS system of the pSR1 plasmid, or any combination thereof.

30. The combination of claim 28, wherein a vector is the plasmid pCXLamIntR.

25 31. The combination of claim 27, wherein a vector is the plasmid pDsRedN1-attB.

32. A kit, comprising the combination of claim 27 and optionally instructions for introducing heterologous nucleic acid into the chromosome.



33. A method for introducing heterologous nucleic acid into a platform artificial chromosome, comprising:

(a) mixing an artificial chromosome comprising at least a first recombination site and a vector comprising at least a second

5 recombination site and the heterologous nucleic acid;

(b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the artificial chromosome.

10 34. The method of claim 33, wherein the artificial chromosome is an ACes.

35. The method of claim 33, wherein said mixing step (a) is conducted in cells ex vivo.

15 36. The method of claim 33, wherein said mixing step (a) is conducted extracellularly in an in vitro reaction mixture.

37. The method of claim 33, wherein the at least one recombination protein is encoded by a bacteriophage selected from the group consisting of bacteriophage lambda, phi 80, P22, P2, 186, P4 and P1.

20 38. The method of claim 37, wherein the at least one recombination protein is encoded by bacteriophage lambda, or mutants thereof.

39. The method of claim 33, wherein at least one recombination protein is selected from the group consisting of Int, IHF, Xis, Cre,  $\gamma\delta$ , Tn3  
25 resolvase, Hin, Gin, Cin and Flp.

40. The method of claim 33, wherein the recombination sites are selected from the group consisting of att and lox P sites.

66. The lambda-intR mutein of claim 65, wherein the lambda-intR mutein comprises SEQ ID NO:37.

67. The method of claim 46 wherein the promoterless marker is transcriptionally downstream of the heterologous nucleic acid, wherein  
5 the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.

68. The method of claim 67, wherein the selectable marker and the heterologous nucleic acid are transcriptionally linked by the presence  
10 of an IRES between them.

69. The method of claim 68, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.

70. The method of claim 69, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase.

71. The method of claim 67 further comprising expressing the  
20 heterologous protein and isolating the heterologous protein.

72. A method for producing a transgenic animal, comprising introducing a platform-ACes into an embryonic cell.

73. The method of claim 72, wherein the embryonic cell is a stem cell.

74. The method of claim 72, wherein the embryonic cell is in an embryo.

75. The method of claim 72, wherein the platform-ACes comprises heterologous nucleic acid that encodes a therapeutic product.

a sequence of nucleotides that targets the vector to an amplifiable region of a chromosome.

92. The vector of claim 91, wherein the amplifiable region comprises heterochromatic nucleic acid.

5 93. The vector of claim 91, wherein the amplifiable region comprises rDNA.

94. The vector of claim 93, wherein the rDNA comprises an intergenic spacer.

95. The vector of claim 91, further comprising nucleic acid  
10 encoding a selectable marker that is not operably associated with any promoter.

96. The vector of claim 91, wherein the chromosome is a mammalian chromosome.

97. The vector of claim 91, wherein the chromosome is a plant  
15 chromosome.

98. A cell of claim 57 that is a plant cell, wherein the ACes platform is a MAC.

99. The plant cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

20 100. The plant cell of claim 99, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

101. A cell of claim 57 that is an animal cell, wherein the ACes platform is a plant artificial chromosome (PAC).

25 102. The cell of claim 101 that is a mammalian cell.

103. The cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

104. The cell of claim 102, wherein the PAC comprises transcriptional regulatory sequence of nucleotides derived from animals.

105. The cell of claim 104, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

106. A method, comprising:

- 5       introducing a vector of claim 91 into a cell;  
      growing the cells; and  
      selecting a cell comprising an artificial chromosome that comprises one or more repeat regions.

107. The method of claim 106, wherein a sufficient portion of the  
10       vector integrates into a chromosome in the cell to result in amplification of chromosomal DNA.

108. The method of claim 106, wherein the artificial chromosome is an *ACes*.

109. A method for screening, comprising:

- 15       contacting a cell comprising a reporter *ACes* with test compounds or known compounds, wherein:

      the reporter *ACes* comprises one or a plurality of reporter constructs;

- a reporter construct comprises a reporter gene in operative linkage  
20       with a regulatory region responsive to test or known compounds; and  
      detecting any increase or decrease in signal output from the reporter, wherein a change in the signal is indicative of activity of the test or known compound on the regulatory region.

110. The method of claim 109, wherein the reporter is operatively  
25       linked to a promoter that controls expression of a gene in a signal transduction pathway, whereby activation or reduction in the signal indicates that the pathway is activated or down-regulated by the test compound.

111. The method of claim 109, wherein the reporter in the construct encodes drug resistance or encodes a fluorescent protein.

112. The method of claim 111, wherein the fluorescent protein is selected from the group consisting of red, green and blue fluorescent  
5 proteins.

113. The method of claim 109, wherein the ACes comprises a plurality of reporter-linked constructs, each with a different reporter, whereby the pathway(s) affected by the test compounds can be elucidated.

10 114. The method of claim 109, wherein a reporter is operatively linked to a promoter that is transcriptionally regulated in response to DNA damage, and the test compounds are genotoxicants.

115. The method of claim 114, wherein the DNA damage is induced by apoptosis, necrosis or cell-cycle perturbations.

15 116. The method of claim 114, wherein unknown compounds are screened to assess whether they are genotoxicants.

117. The method of claim 114, wherein the promoter is a cytochrome P450-profiled promoter.

20 118. The method of claim 114, wherein the cell is in a transgenic animal and toxicity is assessed in the animal.

119. The method of claim 109, wherein:

the cell is a patient cell sample; the patient has a disease;

the regulatory region is one targeted by a drug or drug regimen;

and

25 the method assesses the effectiveness of a treatment for the disease for the particular patient.

120. The method of claim 119, wherein the cell is a tumor cell.

121. The method of claim 109, wherein the cell is a stem cell or a progenitor cell, whereby expression of the reporter is operatively linked to

a regulatory region expressed in the cells to thereby identify stem cells or progenitor cells.

122. The method of claim 109, wherein the cell is in an animal;  
and the method comprises whole-body imaging to monitor expression of  
5 the reporter in the animal.

123. A reporter ACes comprises one or a plurality of reporter constructs, wherein the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

CORRECTED VERSION

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
5 December 2002 (05.12.2002)

PCT

(10) International Publication Number  
WO 02/097059 A3(51) International Patent Classification<sup>7</sup>: C07H 21/04,  
C12N 15/85, 15/87, 15/90, A01N 43/34, C12N 15/09

(21) International Application Number: PCT/US02/17452

(22) International Filing Date: 30 May 2002 (30.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/294,758 30 May 2001 (30.05.2001) US  
60/366,891 21 March 2002 (21.03.2002) US(63) Related by continuation (CON) or continuation-in-part  
(CIP) to earlier applications:US 60/294,758 (CIP)  
Filed on 30 May 2001 (30.05.2001)  
US 60/366,891 (CIP)  
Filed on 21 March 2002 (21.03.2002)(71) Applicant (for all designated States except US): CHROMOSOMOLAR SYSTEMS, INC. [CA/CA]; 8081  
Lougheed Highway, Burnaby, B.C. V5A 1W9 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): PERKINS, Edward  
[US/CA]; 7610 Lawrence Drive, Burnaby, B.C. V5A 1T6  
(CA). PEREZ, Carl [US/CA]; 1201-7680 Granville Avenue,  
Richmond, B.C. V6Y 4B9 (CA). LINDENBAUM, Michael  
[CA/CA]; 252 Finnigan Street, Coquitlam, B.C. V3K 5J7 (CA). GREENE, Amy [US/CA]; 7610 Lawrence  
Drive, Burnaby, B.C. V5A 1T6 (CA). LEUNG, Josephine  
[CA/CA]; 711 Ebert Avenue, Coquitlam, B.C. V3J 7P8  
(CA). FLEMING, Elena [CA/CA]; 248 E 18th, North  
Vancouver, B.C. V7L 2X6 (CA). STEWART, Sandra  
[CA/CA]; 2618 Oxford Street, Vancouver, B.C. V5K 1N3  
(CA). SHELLARD, Joan [CA/CA]; #215-1345 West  
15th Avenue, Vancouver, B.C. V6H 3R3 (CA).(74) Agents: SEIDMAN, Stephanie, L. et al.; Heller Ehrman  
White & McAuliffe LLP, 7th floor, 4350 La Jolla Village  
Drive, San Diego, CA 92122-1246 (US).(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,  
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,  
VN, YU, ZA, ZM, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,  
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent  
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,  
NE, SN, TD, TG).

## Declarations under Rule 4.17:

— as to the applicant's entitlement to claim the priority of the  
earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY,  
BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC,  
EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN,  
IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV,  
MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH,  
PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR,  
TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent  
(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB,  
GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ,  
CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,  
TG)— as to the applicant's entitlement to claim the priority of the  
earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY,  
BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC,  
EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN,  
IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV,  
MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH,  
PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR,  
TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent  
(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB,

[Continued on next page]

(54) Title: CHROMOSOME-BASED PLATFORMS

(57) Abstract: Artificial chromosomes, including *Aces*, that have been engineered to contain available sites for site-specific, recombination-directed integration of DNA of interest are provided. These artificial chromosomes permit tractable, efficient, rational engineering of the chromosome for a variety of applications.

WO 02/097059 A3



GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

**Published:**

— with international search report

**(88) Date of publication of the international search report:**  
30 May 2003

**(48) Date of publication of this corrected version:**

25 September 2003

**(15) Information about Correction:**

see PCT Gazette No. 39/2003 of 25 September 2003, Section II

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



-1-

**CHROMOSOME-BASED PLATFORMS****RELATED APPLICATIONS**

Benefit of priority to U.S. provisional application Serial No. 60/294,758, filed May 30, 2001, to Perkins, *et al.*, entitled "CHROMOSOME-BASED PLATFORMS" and to U.S. provisional application Serial No. 60/366,891, filed March 21, 2002, to Perkins, *et al.*, entitled

5 "CHROMOSOME-BASED PLATFORMS" is claimed. Where permitted, the subject matter of which are herein incorporated by reference in their entirety.

This application is related to Provisional Application No. 60/294,687, filed May 30, 2001, by CARL PEREZ AND STEVEN

10 FABIJANSKI entitled *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING PLANT ARTIFICIAL CHROMOSOMES* and to U.S. Provisional Application No. 60/296,329, filed June 4, 2001, by CARL PEREZ AND STEVEN FABIJANSKI entitled

15 *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING PLANT ARTIFICIAL CHROMOSOMES*. This application also is related to U.S. Provisional Application No. 60/294,758, filed May 30, 2001, by EDWARD PERKINS *et al.*, entitled *CHROMOSOME-BASED PLATFORMS* and to U.S. Provisional Application No. 60/366,891, filed

20 March 21, 2002, by EDWARD PERKINS *et al.*, entitled *CHROMOSOME-BASED PLATFORMS*. This application is also related to U.S. application Serial Nos. (attorney dkt nos. 24601-419 and 419PC), filed on the same day herewith, entitled *PLANT ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS OF PREPARING PLANT ARTIFICIAL CHROMOSOMES* to Perez *et al.*.

25 This application is related to U.S. application Serial No. 08/695,191, filed August 7, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND*

*METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*, now U.S. Patent No. 6,025,155. This application is also related to U.S. application Serial No. 08/682,080, filed July 15, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF*  
5 *AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*, now U.S. Patent No. 6,077,697. This application is also related U.S. application Serial No. 08/629,822, filed April 10, 1996 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING*  
10 *ARTIFICIAL CHROMOSOMES* (now abandoned), and is also related to copending U.S. application Serial No. 09/096,648, filed June 12, 1998, by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* and to U.S. application Serial No.  
15 09/835,682, April 10, 1997 by GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES* (now abandoned). This application is also related to copending U.S. application Serial No. 09/724,726, filed November 28, 2000, U.S. application Serial  
20 No. 09/724,872, filed November 28, 2000, U.S. application Serial No. 09/724,693, filed November 28, 2000, U.S. application Serial No. 09/799,462, filed March 5, 2001, U.S. application Serial No. 09/836,911, filed April 17, 2001, and U.S. application Serial No. 10/125,767, filed April 17, 2002, each of which is by GYULA  
25 HADLACZKY and ALADAR SZALAY, and is entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*. This application is also related to International PCT application No. WO 97/40183. Where permitted the

subject matter of each of these provisional applications, international applications, and applications is incorporated by reference in its entirety.

#### **FIELD OF INVENTION**

Artificial chromosomes, including *ACes*, that have been engineered  
5 to contain available sites for site-specific, recombination-directed  
integration of DNA of interest are provided. These artificial chromosomes  
permit tractable, efficient, rational engineering of the chromosome.

#### **BACKGROUND**

##### **Artificial chromosomes**

10 A variety of artificial chromosomes for use in plants and animals,  
particularly higher plants and animals are available. In particular, U.S.  
Patent Nos. 6,025,155 and 6,077,697 provide heterochromatic artificial  
chromosomes designated therein as satellite artificial chromosomes  
(SATACs) and now designated artificial chromosome expression systems  
15 (*ACes*). These chromosomes are prepared by introducing heterologous  
DNA into a selected plant or animal cell under conditions that result in  
integration into a region of the chromosome that leads to an amplification  
event resulting in production of a dicentric chromosome. Subsequent  
treatment and growth of cells with dicentric chromosomes, including  
20 further amplifications, ultimately results in the artificial chromosomes  
provided therein. In order to introduce a desired heterologous gene (or a  
plurality of heterologous genes) into the artificial chromosome, the  
process is repeated introducing the desired heterologous genes and  
nucleic acids in the initial targeting step. This process is time consuming  
25 and tedious. Hence, more tractable and efficient methods for introducing  
heterologous nucleic acid molecules into artificial chromosomes,  
particularly *ACes*, are needed.

Therefore, it is an object herein to provide engineered artificial chromosomes that permit tractable, efficient and rational engineering of artificial chromosomes.

#### **SUMMARY OF THE INVENTION**

5            Provided herein are artificial chromosomes that permit tractable, efficient and rational engineering thereof. In particular, the artificial chromosomes provided herein contain one or a plurality of loci (sites) for site-specific, recombination-directed integration of DNA. Thus, provided  
10            herein are platform artificial chromosome expression systems ("platform *ACes*") containing single or multiple site-specific, recombination sites. The artificial chromosomes and *ACes* artificial chromosomes include plant and animal chromosomes. Any recombinase system that effects site-specific recombination is contemplated for use herein.

             In one embodiment, chromosomes, including platform *ACes*, are  
15            provided that contain one or more lambda *att* sites designed for recombination-directed integration in the presence of lambda integrase, and that are mutated so that they do not require additional factors. Methods for preparing such chromosomes, vectors for use in the methods, and uses of the resulting chromosomes are also provided.

20            Platform *ACes* containing the recombination site(s) and methods for introducing heterologous nucleic acid into such sites and vectors therefor, are provided.

             Also provided herein is a bacteriophage lambda ( $\lambda$ ) integrase site-specific recombination system.

25            Methods using recombinase mediated recombination target gene expression vectors and/or genes for insertion thereof into platform chromosomes and the resulting chromosomes are provided.

             Combinations and kits containing the combinations of vectors encoding a recombinase and integrase and primers for introduction of the

site recognized thereby are also provided. The kits optionally include instructions for performing site-directed integration or preparation of *ACes* containing such sites.

Also provided herein are mammalian and plant cells comprising the  
5 artificial chromosomes and *ACes* described herein. The cells can be nuclear donor cells, stem cells, such as a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell or an embryonic stem cell.

Also provided is a lambda-intR mutein comprising a glutamic acid to arginine change at position 174 of wild-type lambda-integrase3. Also  
10 provided are transgenic animals and methods for producing a transgenic animal, comprising introducing a *ACes* into an embryonic cell, such as a stem cell or embryo. The *ACes* can comprise heterologous nucleic acid that encodes a therapeutic product. The transgenic animal can be a fish, insect, reptile, amphibians, arachnid or mammal. In certain embodiments,  
15 the *ACes* is introduced by cell fusion, lipid-mediated transfection by a carrier system, microinjection, microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer.

The platform *ACes*, including plant and animal *ACes*, such as MACs, provided herein can be introduced into cells, such as, but not  
20 limited to, animal cells, including mammalian cells, and into plant cells. Hence plant cells that contain platform MACs, animal cells that contain platform PACs and other combinations of cells and platform *ACes* are provided.

#### **DESCRIPTION OF FIGURES**

25 FIGURE 1 provides a diagram depicting creation of an exemplary *ACes* artificial chromosome prepared using methods detailed in U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183. In this exemplified embodiment, the nucleic acid is targeted to an acrocentric chromosome in an animal or plant, and the

heterologous nucleic acid includes a sequence-specific recombination site and marker genes.

FIGURE 2 provides a map of pWEPuro9K, which is a targeting vector derived from the vector pWE15 (GenBank Accession # X65279; SEQ ID No. 31). Plasmid pWE15 was modified by replacing the *SaI* (Klenow filled)/*SmaI* neomycin resistance encoding fragment with the *PvuII*/*Bam*HI (Klenow filled) puromycin resistance-encoding fragment (isolated from plasmid pPUR, Clontech Laboratories, Inc., Palo Alto, CA; GenBank Accession no. U07648; SEQ ID No. 30) resulting in plasmid pWEPuro. Subsequently a 9 Kb *NotI* fragment from the plasmid pFK161 (see Example 1, see, also Csonka *et al.* (2000) *Journal of Cell Science* 113:3207-32161; and SEQ ID NO: 118), containing a portion of the mouse rDNA region, was cloned into the *NotI* site of pWEPuro resulting in plasmid pWEPuro9K.

FIGURE 3 depicts construction of an ACes platform chromosome with a single recombination site, such as loxP sites or an *attP* or *attB* site. This platform ACes chromosome is an exemplary artificial chromosome with a single recombination site.

FIGURE 4 provides a map of plasmid pSV40-193attPsensePur.

FIGURE 5 depicts a method for formation of a chromosome platform with multiple recombination integration sites, such as *attP* sites.

FIGURE 6 sets forth the sequences of the core region of *attP*, *attB*, *attL* and *attR* (SEQ ID Nos. 33-36).

FIGURE 7 depicts insertional recombination of a vector encoding a marker gene, DsRed and an *attB* site with an artificial chromosome containing an *attP* site.

FIGURE 8 provides a map of plasmid pCXLamIntR (SEQ ID NO: 112), which includes the Lambda integrase (E174R)-encoding nucleic acid.

FIGURE 9 diagrammatically summarizes the platform technology; marker 1 permits selection of the artificial chromosomes containing the integration site; marker 2, which is promoterless in the target gene expression vector, permits selection of recombinants. Upon  
5 recombination with the platform marker 2 is expressed under the control of a promoter resident on the platform.

FIGURE 10 provides the vector map for the plasmid p18attBZEO-5'6XHS4eGFP (SEQ ID NO: 116).

FIGURE 11 provides the vector map for the plasmid p18attBZEO-  
10 3'6XHS4eGFP (SEQ ID NO: 115).

FIGURE 12 provides the vector map for the plasmid p18attBZEO-(6XHS4)2eGFP (SEQ ID NO: 110).

FIGURES 13 AND 14 depict the integration of a PCR product by site-specific recombination as set forth in Example 8.

15 FIGURE 15 provides the vector map for the plasmid pPACrDNA as set forth in Example 9.A.

## **DETAILED DESCRIPTION OF THE INVENTION**

### **A. DEFINITIONS**

Unless defined otherwise, all technical and scientific terms used  
20 herein have the same meaning as is commonly understood by one of skill in the art to which the invention(s) belong. All patents, patent applications, published applications and publications, Genbank sequences, websites and other published materials referred to throughout the entire disclosure herein, unless noted otherwise, are incorporated by reference  
25 in their entirety. Where reference is made to a URL or other such identifier or address, it understood that such identifiers can change and particular information on the internet can come and go, but equivalent information can be found by searching the internet. Reference thereto evidences the availability and public dissemination of such information.

As used herein, nucleic acid refers to single-stranded and/or double-stranded polynucleotides, such as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), as well as analogs or derivatives of either RNA or DNA. Also included in the term "nucleic acid" are analogs of nucleic acids such as peptide nucleic acid (PNA), phosphorothioate DNA, and other such analogs and derivatives. When referring to probes or primers, optionally labeled, with a detectable label, such as a fluorescent or radiolabel, single-stranded molecules are contemplated. Such molecules are typically of a length such that they are statistically unique and of low copy number (typically less than 5, preferably less than 3) for probing or priming a library. Generally a probe or primer contains at least 14, 16 or 30 contiguous nucleotides of sequence complementary to or identical to a gene of interest. Probes and primers can be 10, 20, 30, 50, 100 or more nucleotides long.

As used herein, DNA is meant to include all types and sizes of DNA molecules including cDNA, plasmids and DNA including modified nucleotides and nucleotide analogs.

As used herein, nucleotides include nucleoside mono-, di-, and triphosphates. Nucleotides also include modified-nucleotides, such as, but are not limited to, phosphorothioate nucleotides and deazapurine nucleotides and other nucleotide analogs.

As used herein, heterologous or foreign DNA and RNA are used interchangeably and refer to DNA or RNA that does not occur naturally as part of the genome in which it is present or which is found in a location or locations and/or in amounts in a genome or cell that differ from that in which it occurs in nature. Heterologous nucleic acid is generally not endogenous to the cell into which it is introduced, but has been obtained from another cell or prepared synthetically. Generally, although not necessarily, such nucleic acid encodes RNA and proteins that are not



normally produced by the cell in which it is expressed. Any DNA or RNA that one of skill in the art would recognize or consider as heterologous or foreign to the cell in which it is expressed is herein encompassed by heterologous DNA. Heterologous DNA and RNA may also encode RNA or proteins that mediate or alter expression of endogenous DNA by affecting transcription, translation, or other regulatable biochemical processes.

Examples of heterologous DNA include, but are not limited to, DNA that encodes a gene product or gene product(s) of interest, introduced for purposes of modification of the endogenous genes or for production of an encoded protein. For example, a heterologous or foreign gene may be isolated from a different species than that of the host genome, or alternatively, may be isolated from the host genome but operably linked to one or more regulatory regions which differ from those found in the unaltered, native gene. Other examples of heterologous DNA include, but are not limited to, DNA that encodes traceable marker proteins, such as a protein that confers traits including, but not limited to, herbicide, insect, or disease resistance; traits, including, but not limited to, oil quality or carbohydrate composition. Antibodies that are encoded by heterologous DNA may be secreted or expressed on the surface of the cell in which the heterologous DNA has been introduced.

As used herein, operative linkage or operative association, or grammatical variations thereof, of heterologous DNA to regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences refers to the relationship between such DNA and such sequences of nucleotides. For example, operative linkage of heterologous DNA to a promoter refers to the physical relationship between the DNA and the promoter such that the transcription of such DNA is initiated from the

promoter by an RNA polymerase that specifically recognizes, binds to and transcribes the DNA.

In order to optimize expression and/or *in vitro* transcription, it may be necessary to remove, add or alter 5' untranslated portions of the clones to eliminate extra, potential inappropriate alternative translation initiation (*i.e.*, start) codons or other sequences that may interfere with or reduce expression, either at the level of transcription or translation. Alternatively, consensus ribosome binding sites (see, *e.g.*, Kozak (1991) *J. Biol. Chem.* 266:19867-19870) can be inserted immediately 5' of the start codon and may enhance expression.

As used herein, a sequence complementary to at least a portion of an RNA, with reference to antisense oligonucleotides, means a sequence having sufficient complementarity to be able to hybridize with the RNA, preferably under moderate or high stringency conditions, forming a stable duplex. The ability to hybridize depends on the degree of complementarity and the length of the antisense nucleic acid. The longer the hybridizing nucleic acid, the more base mismatches it can contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

As used herein, regulatory molecule refers to a polymer of deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) or a polypeptide that is capable of enhancing or inhibiting expression of a gene.

As used herein, recognition sequences are particular sequences of nucleotides that a protein, DNA, or RNA molecule, or combinations thereof, (such as, but not limited to, a restriction endonuclease, a modification methylase and a recombinase) recognizes and binds. For example, a recognition sequence for Cre recombinase (see, *e.g.*, SEQ ID NO:58) is a 34 base pair sequence containing two 13 base pair inverted

repeats (serving as the recombinase binding sites) flanking an 8 base pair core and designated loxP (see, *e.g.*, Sauer (1994) *Current Opinion in Biotechnology* 5:521-527). Other examples of recognition sequences, include, but are not limited to, attB and attP, attR and attL and others  
5 (see, *e.g.*, SEQ ID Nos. 8, 41-56 and 72), that are recognized by the recombinase enzyme Integrase (see, SEQ ID Nos. 37 and 38 for the nucleotide and encoded amino acid sequences of an exemplary lambda phage integrase).

The recombination site designated attB is an approximately 33 base  
10 pair sequence containing two 9 base pair core-type Int binding sites and a 7 base pair overlap region; attP (SEQ ID No. 72) is an approximately 240 base pair sequence containing core-type Int binding sites and arm-type Int binding sites as well as sites for auxiliary proteins IHF, FIS, and Xis (see, *e.g.*, Landy (1993) *Current Opinion in Biotechnology* 3:699-707) see,  
15 *e.g.*, SEQ ID Nos. 8 and 72).

As used herein, a recombinase is an enzyme that catalyzes the exchange of DNA segments at specific recombination sites. An integrase herein refers to a recombinase that is a member of the lambda ( $\lambda$ ) integrase family.

20 As used herein, recombination proteins include excisive proteins, integrative proteins, enzymes, co-factors and associated proteins that are involved in recombination reactions using one or more recombination sites (see, Landy (1993) *Current Opinion in Biotechnology* 3:699-707). The recombination proteins used herein can be delivered to a cell via an  
25 expression cassette on an appropriate vector, such as a plasmid, and the like. In other embodiments, the recombination proteins can be delivered to a cell in protein form in the same reaction mixture used to deliver the desired nucleic acid, such as a platform ACes, donor target vectors, and the like.

As used herein the expression "lox site" means a sequence of nucleotides at which the gene product of the cre gene, referred to herein as Cre, can catalyze a site-specific recombination event. A LoxP site is a 34 base pair nucleotide sequence from bacteriophage P1 (see, 5 *e.g.*, Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398-3402). The LoxP site contains two 13 base pair inverted repeats separated by an 8 base pair spacer region as follows: (SEQ ID NO. 57):

ATAACTTCGTATA ATGTATGC TATACGAAGTTAT

*E. coli*/DH5Δlac and yeast strain BSY23 transformed with plasmid pBS44 10 carrying two loxP sites connected with a LEU2 gene are available from the American Type Culture Collection (ATCC) under accession numbers ATCC 53254 and ATCC 20773, respectively. The lox sites can be isolated from plasmid pBS44 with restriction enzymes *EcoRI* and *SalI*, or *XhoI* and *BamHI*. In addition, a preselected DNA segment can be inserted 15 into pBS44 at either the *SalI* or *BamHI* restriction enzyme sites. Other lox sites include, but are not limited to, LoxB, LoxL, LoxC2 and LoxR sites, which are nucleotide sequences isolated from *E. coli* (see, *e.g.*, Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398). Lox sites can also be produced by a variety of synthetic techniques (see, *e.g.*, Ito *et al.* (1982) 20 *Nuc. Acid Res.* 10:1755 and Ogilvie *et al.* (1981) *Science* 270:270).

As used herein, the expression "cre gene" means a sequence of nucleotides that encodes a gene product that effects site-specific recombination of DNA in eukaryotic cells at lox sites. One cre gene can be isolated from bacteriophage P1 (see, *e.g.*, Abremski *et al.* (1983) *Cell* 25 32:1301-1311). *E. coli* DH1 and yeast strain BSY90 transformed with plasmid pBS39 carrying a cre gene isolated from bacteriophage P1 and a GAL1 regulatory nucleotide sequence are available from the American Type Culture Collection (ATCC) under accession numbers ATCC 53255

and ATCC 20772, respectively. The cre gene can be isolated from plasmid pBS39 with restriction enzymes *XhoI* and *SaII*.

As used herein, site-specific recombination refers to site-specific recombination that is effected between two specific sites on a single  
5 nucleic acid molecule or between two different molecules that requires the presence of an exogenous protein, such as an integrase or recombinase.

For example, Cre-lox site-specific recombination can include the following three events:

- 10 a. deletion of a pre-selected DNA segment flanked by lox sites;
- b. inversion of the nucleotide sequence of a pre-selected DNA segment flanked by lox sites; and
- c. reciprocal exchange of DNA segments proximate to  
15 lox sites located on different DNA molecules.

This reciprocal exchange of DNA segments can result in an integration event if one or both of the DNA molecules are circular. DNA segment refers to a linear fragment of single- or double-stranded deoxyribonucleic acid (DNA), which can be derived from any source.

20 Since the lox site is an asymmetrical nucleotide sequence, two lox sites on the same DNA molecule can have the same or opposite orientations with respect to each other. Recombination between lox sites in the same orientation results in a deletion of the DNA segment located between the two lox sites and a connection between the resulting ends of the original  
25 DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule result in an inversion of the nucleotide sequence of the DNA segment located between the two lox

sites. In addition, reciprocal exchange of DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the gene product of the cre gene. Thus, the Cre-lox system can be used to specifically delete, invert, or  
5 insert DNA. The precise event is controlled by the orientation of lox DNA sequences, in *cis* the lox sequences direct the *Cre* recombinase to either delete (lox sequences in direct orientation) or invert (lox sequences in inverted orientation) DNA flanked by the sequences, while *in trans* the lox sequences can direct a homologous recombination event resulting in the  
10 insertion of a recombinant DNA.

As used herein, a chromosome is a nucleic acid molecule, and associated proteins, that is capable of replication and segregation within a cell upon cell division. Typically, a chromosome contains a centromeric region, replication origins, telomeric regions and a region of nucleic acid  
15 between the centromeric and telomeric regions.

As used herein, a centromere is any nucleic acid sequence that confers an ability to segregate to daughter cells through cell division. A centromere may confer stable segregation of a nucleic acid sequence, including an artificial chromosome containing the centromere, through  
20 mitotic or meiotic divisions, including through both mitotic and meiotic divisions. A particular centromere is not necessarily derived from the same species in which it is introduced, but has the ability to promote DNA segregation in cells of that species.

As used herein, euchromatin and heterochromatin have their  
25 recognized meanings. Euchromatin refers to chromatin that stains diffusely and that typically contains genes, and heterochromatin refers to chromatin that remains unusually condensed and that has been thought to be transcriptionally inactive. Highly repetitive DNA sequences (satellite DNA) are usually located in regions of the heterochromatin surrounding

the centromere (pericentric or pericentromeric heterochromatin).

Constitutive heterochromatin refers to heterochromatin that contains the highly repetitive DNA which is constitutively condensed and genetically inactive.

5           As used herein, an acrocentric chromosome refers to a chromosome with arms of unequal length.

          As used herein, endogenous chromosomes refer to genomic chromosomes as found in a cell prior to generation or introduction of an artificial chromosome.

10           As used herein, artificial chromosomes are nucleic acid molecules, typically DNA, that stably replicate and segregate alongside endogenous chromosomes in cells and have the capacity to accommodate and express heterologous genes contained therein. It has the capacity to act as a gene delivery vehicle by accommodating and expressing foreign genes  
15           contained therein. A mammalian artificial chromosome (MAC) refers to chromosomes that have an active mammalian centromere(s). Plant artificial chromosomes, insect artificial chromosomes and avian artificial chromosomes refer to chromosomes that include centromeres that function in plant, insect and avian cells, respectively. A human artificial  
20           chromosome (HAC) refers to chromosomes that include centromeres that function in human cells. For exemplary artificial chromosomes, see, *e.g.*, U.S. Patent Nos. 6,025,155; 6,077,697; 5,288,625; 5,712,134; 5,695,967; 5,869,294; 5,891,691 and 5,721,118 and published International PCT application Nos, WO 97/40183 and WO 98/08964.  
25           Artificial chromosomes include those that are predominantly heterochromatic (formerly referred to as satellite artificial chromosomes (SATACs); see, *e.g.*, U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183), minichromosomes that contain a *de novo* centromere (see, U.S. Patent

Nos. 5,712,134, 5,891,691 and 5,288,625), artificial chromosomes predominantly made up of repeating nucleic acid units and that contain substantially equivalent amounts of euchromatic and heterochromatic DNA and *in vitro* assembled artificial chromosomes (see, copending U.S. 5 provisional application Serial No. 60/294,687, filed on May 30, 2001).

As used herein, the term "satellite DNA-based artificial chromosome (SATAC)" is interchangeable with the term "artificial chromosome expression system (ACes)". These artificial chromosomes (ACes) include those that are substantially all neutral non-coding 10 sequences (heterochromatin) except for foreign heterologous, typically gene-encoding nucleic acid, that is interspersed within the heterochromatin for the expression therein (see U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183), or that is in a single locus as provided herein. Also included 15 are ACes that may include euchromatin and that result from the process described in U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183 and outlined herein. The delineating structural feature is the presence of repeating units, that are generally predominantly heterochromatin. The precise structure of the ACes will 20 depend upon the structure of the chromosome in which the initial amplification event occurs; all share the common feature of including a defined pattern of repeating units. Generally ACes have more heterochromatin than euchromatin. Foreign nucleic acid molecules (heterologous genes) contained in these artificial chromosome expression 25 systems can include any nucleic acid whose expression is of interest in a particular host cell. Such foreign nucleic acid molecules, include, but are not limited to, nucleic acid that encodes traceable marker proteins (reporter genes), such as fluorescent proteins, such as green, blue or red fluorescent proteins (GFP, BFP and RFP, respectively), other reporter



genes, such as  $\beta$ -galactosidase and proteins that confer drug resistance, such as a gene encoding hygromycin-resistance. Other examples of heterologous nucleic acid molecules include, but are not limited to, DNA that encodes therapeutically effective substances, such as anti-cancer  
5 agents, enzymes and hormones, DNA that encodes other types of proteins, such as antibodies, and DNA that encodes RNA molecules (such as antisense or siRNA molecules) that are not translated into proteins.

As used herein, an artificial chromosome platform, also referred to  
10 herein as a "platform ACes" or "ACes platform", refers to an artificial chromosome that has been engineered to include one or more sites for site-specific, recombination-directed integration. In particular, ACes that are so-engineered are provided. Any sites, including but not limited to any described herein, that are suitable for such integration are  
15 contemplated. Plant and animal platform ACes are provided. Among the ACes contemplated herein are those that are predominantly heterochromatic (formerly referred to as satellite artificial chromosomes (SATACs); see, *e.g.*, U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183), artificial  
20 chromosomes predominantly made up of repeating nucleic acid units and that contain substantially equivalent amounts of euchromatic and heterochromatic DNA resulting from an amplification event depicted in the referenced patent and herein. Included among the ACes for use in generating platforms, are artificial chromosomes that introduce and  
25 express heterologous nucleic acids in plants (see, copending U.S. provisional application Serial No. 60/294,687, filed on May 30, 2001). These include artificial chromosomes that have a centromere derived from a plant, and, also, artificial chromosomes that have centromeres that may be derived from other organisms but that function in plants.

As used herein a "reporter ACes" refers to an ACes that comprises one or a plurality of reporter constructs, where the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.

5 As used herein, amplification, with reference to DNA, is a process in which segments of DNA are duplicated to yield two or multiple copies of substantially similar or identical or nearly identical DNA segments that are typically joined as substantially tandem or successive repeats or inverted repeats.

10 As used herein, amplification-based artificial chromosomes are artificial chromosomes derived from natural or endogenous chromosomes by virtue of an amplification event, such as one initiated by introduction of heterologous nucleic acid into rDNA in a chromosome. As a result of such an event, chromosomes and fragments thereof exhibiting segmented or repeating patterns arise. Artificial chromosomes can be formed from these chromosomes and fragments. Hence, amplification-based artificial chromosomes refer to engineered chromosomes that exhibit an ordered segmentation that is not observed in naturally occurring chromosomes and that distinguishes them from naturally occurring chromosomes. The  
15 segmentation, which can be visualized using a variety of chromosome analysis techniques known to those of skill in the art, correlates with the structure of these artificial chromosomes. In addition to containing one or more centromeres, the amplification-based artificial chromosomes, throughout the region or regions of segmentation are predominantly made  
20 up of nucleic acid units also referred to as "amplicons", that is (are) repeated in the region and that have a similar gross structure. Repeats of an amplicon tend to be of similar size and share some common nucleic acid sequences. For example, each repeat of an amplicon may contain a replication site involved in amplification of chromosome segments and/or  
25

some heterologous nucleic acid that was utilized in the initial production of the artificial chromosome. Typically, the repeating units are substantially similar in nucleic acid composition and may be nearly identical.

5           The amplification-based artificial chromosomes differ depending on the chromosomal region that has undergone amplification in the process of artificial chromosome formation. The structures of the resulting chromosomes can vary depending upon the initiating event and/or the conditions under which the heterologous nucleic acid is introduced,  
10 including modification to the endogenous chromosomes. For example, in some of the artificial chromosomes provided herein, the region or regions of segmentation may be made up predominantly of heterochromatic DNA. In other artificial chromosomes provided herein, the region or regions of segmentation may be made up predominantly of euchromatic DNA or may  
15 be made up of similar amounts of heterochromatic and euchromatic DNA.

As used herein an amplicon is a repeated nucleic acid unit. In some of the artificial chromosomes described herein, an amplicon may contain a set of inverted repeats of a megareplicon. A megareplicon represents a higher order replication unit. For example, with reference to  
20 some of the predominantly heterochromatic artificial chromosomes, the megareplicon can contain a set of tandem DNA blocks (*e.g.*, ~7.5 Mb DNA blocks) each containing satellite DNA flanked by non-satellite DNA or may be made up of substantially rDNA. Contained within the megareplicon is a primary replication site, referred to as the  
25 megareplicator, which may be involved in organizing and facilitating replication of the pericentric heterochromatin and possibly the centromeres. Within the megareplicon there may be smaller (*e.g.*, 50-300 kb) secondary replicons.

In artificial chromosomes, such as those provided U.S. Patent Nos. 6,025,155 and 6,077,697 and International PCT application No. WO 97/40183, the megareplicon is defined by two tandem blocks (~7.5 Mb DNA blocks in the chromosomes provided therein). Within each artificial chromosome or among a population thereof, each amplicon has the same gross structure but may contain sequence variations. Such variations will arise as a result of movement of mobile genetic elements, deletions or insertions or mutations that arise, particularly in culture. Such variation does not affect the use of the artificial chromosomes or their overall structure as described herein.

As used herein, amplifiable, when used in reference to a chromosome, particularly the method of generating artificial chromosomes provided herein, refers to a region of a chromosome that is prone to amplification. Amplification typically occurs during replication and other cellular events involving recombination (*e.g.*, DNA repair). Such regions include regions of the chromosome that contain tandem repeats, such as satellite DNA, rDNA, and other such sequences.

As used herein, a dicentric chromosome is a chromosome that contains two centromeres. A multicentric chromosome contains more than two centromeres.

As used herein, a formerly dicentric chromosome is a chromosome that is produced when a dicentric chromosome fragments and acquires new telomeres so that two chromosomes, each having one of the centromeres, are produced. Each of the fragments is a replicable chromosome. If one of the chromosomes undergoes amplification of primarily euchromatic DNA to produce a fully functional chromosome that is predominantly (at least more than 50%) euchromatin, it is a minichromosome. The remaining chromosome is a formerly dicentric chromosome. If one of the chromosomes undergoes amplification,

whereby heterochromatin (such as, for example, satellite DNA) is amplified and a euchromatic portion (such as, for example, an arm) remains, it is referred to as a sausage chromosome. A chromosome that is substantially all heterochromatin, except for portions of heterologous DNA, is called a predominantly heterochromatic artificial chromosome. Predominantly heterochromatic artificial chromosomes can be produced from other partially heterochromatic artificial chromosomes by culturing the cell containing such chromosomes under conditions such as BrdU treatment that destabilize the chromosome and/or growth under selective conditions so that a predominantly heterochromatic artificial chromosome is produced. For purposes herein, it is understood that the artificial chromosomes may not necessarily be produced in multiple steps, but may appear after the initial introduction of the heterologous DNA. Typically, artificial chromosomes appear after about 5 to about 60, or about 5 to about 55, or about 10 to about 55 or about 25 to about 55 or about 35 to about 55 cell doublings after initiation of artificial chromosome generation, or they may appear after several cycles of growth under selective conditions and BrdU treatment.

As used herein, an artificial chromosome that is predominantly heterochromatic (*i.e.*, containing more heterochromatin than euchromatin, typically more than about 50%, more than about 70%, or more than about 90% heterochromatin) may be produced by introducing nucleic acid molecules into cells, such as, for example, animal or plant cells, and selecting cells that contain a predominantly heterochromatic artificial chromosome. Any nucleic acid may be introduced into cells in such methods of producing the artificial chromosomes. For example, the nucleic acid may contain a selectable marker and/or optionally a sequence that targets nucleic acid to the pericentric, heterochromatic region of a chromosome, such as in the short arm of acrocentric chromosomes and

nucleolar organizing regions. Targeting sequences include, but are not limited to, lambda phage DNA and rDNA for production of predominantly heterochromatic artificial chromosomes in eukaryotic cells.

After introducing the nucleic acid into cells, a cell containing a  
5 predominantly heterochromatic artificial chromosome is selected. Such  
cells may be identified using a variety of procedures. For example,  
repeating units of heterochromatic DNA of these chromosomes may be  
discerned by G-banding and/or fluorescence in situ hybridization (FISH)  
techniques. Prior to such analyses, the cells to be analyzed may be  
10 enriched with artificial chromosome-containing cells by sorting the cells  
on the basis of the presence of a selectable marker, such as a reporter  
protein, or by growing (culturing) the cells under selective conditions. It  
is also possible, after introduction of nucleic acids into cells, to select  
cells that have a multicentric, typically dicentric, chromosome, a formerly  
15 multicentric (typically dicentric) chromosome and/or various  
heterochromatic structures, such as a megachromosome and a sausage  
chromosome, that contain a centromere and are predominantly  
heterochromatic and to treat them such that desired artificial  
chromosomes are produced. Cells containing a new chromosome are  
20 selected. Conditions for generation of a desired structure include, but are  
not limited to, further growth under selective conditions, introduction of  
additional nucleic acid molecules and/or growth under selective conditions  
and treatment with destabilizing agents, and other such methods (see  
International PCT application No. WO 97/40183 and U.S. Patent Nos.  
25 6,025,155 and 6,077,697).

As used herein, a "selectable marker" is a nucleic acid segment,  
generally DNA, that allows one to select for or against a molecule or a  
cell that contains it, often under particular conditions. These markers can  
encode an activity, such as, but not limited to, production of RNA,

peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds and compositions. Examples of selectable markers include but are not limited to: (1) nucleic acid segments that encode products that provide resistance against otherwise  
5 toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid segments that encode products that can be identified, such as phenotypic markers,  
10 including  $\beta$ -galactosidase, red, blue and/or green fluorescent proteins (FPs), and cell surface proteins; (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described in Nos. 1-5 above (e.g., antisense  
15 oligonucleotides or siRNA molecules for use in RNA interference); (7) nucleic acid segments that bind products that modify a substrate (e.g. restriction endonucleases); (8) nucleic acid segments that can be used to isolate a desired molecule (e.g. specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence that can be  
20 otherwise non-functional, such as for PCR amplification of subpopulations of molecules; and/or (10) nucleic acid segments, which when absent, directly or indirectly confer sensitivity to particular compounds. Thus, for example, selectable markers include nucleic acids encoding fluorescent proteins, such as green fluorescent proteins,  $\beta$ -galactosidase and other  
25 readily detectable proteins, such as chromogenic proteins or proteins capable of being bound by an antibody and FACs sorted. Selectable markers such as these, which are not required for cell survival and/or proliferation in the presence of a selection agent, are also referred to herein as reporter molecules. Other selectable markers, e.g., the

neomycin phosphotransferase gene, provide for isolation and identification of cells containing them by conferring properties on the cells that make them resistant to an agent, *e.g.*, a drug such as an antibiotic, that inhibits proliferation of cells that do not contain the marker.

5           As another example, interference of gene expression by double stranded RNA has been shown in *Caenorhabditis elegans*, plants, *Drosophila*, protozoans and mammals. This method is known as RNA interference (RNAi) and utilizes short, double-stranded RNA molecules (siRNAs). The siRNAs are generally composed of a 19-22bp double-  
10           stranded RNA stem, a loop region and a 1-4 bp overhang on the 3' end. The reduction of gene expression has been accomplished by direct introduction of the siRNAs into the cell (Harborth J et al., 2001, J Cell Sci 114(pt 24):4557-65) as well as the introduction of DNA encoding and expressing the siRNA molecule. The encoded siRNA molecules are under  
15           the regulation of an RNA polymerase III promoter (see, *e.g.*, Yu et al., 2002, Proc Natl Acad Sci USA 99(9):6047-52; Brummelkamp et al., 2002, Science 296(5567):550-3; Miyagishi et al., 2002, Nat Biotechnol 20(5):497-500; and the like). In certain embodiments, RNAi in mammalian cells may have advantages over other therapeutic methods.  
20           For example, producing siRNA molecules that block viral genetic activities in infected cells may reduce the effects of the virus. Platform ACes provided herein encoding siRNA molecule(s) are an additional utilization of the platform ACes technology. The platform ACes could be engineered to encode one or more siRNA molecules to create gene "knockdowns". In  
25           one embodiment, a platform ACes can be engineered to encode both the siRNA molecule and a replacement gene. For example, a mouse model or cell culture system could be generated using a platform ACes that has a knockdown of the endogenous mouse gene, by siRNA, and the human gene homolog expressing in place of the mouse gene. The placement of



siRNA encoding sequences under the regulation of a regulatable or inducible promoter would allow one to temporally and/or spatially control the knockdown effect of the corresponding gene.

As used herein, a reporter gene includes any gene that expresses a detectable gene product, which may be RNA or protein. Generally  
5 reporter genes are readily detectable. Examples of reporter genes include, but are not limited to nucleic acid encoding a fluorescent protein, CAT (chloramphenicol acetyl transferase) (Alton *et al.* (1979) *Nature* 282: 864-869) luciferase, and other enzyme detection systems, such as beta-  
10 galactosidase; firefly luciferase (deWet *et al.* (1987) *Mol. Cell. Biol.* 7:725-737); bacterial luciferase (Engebrecht and Silverman (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81:4154-4158; Baldwin *et al.* (1984) *Biochemistry* 23:3663-3667); and alkaline phosphatase (Toh *et al.* (1989) *Eur. J. Biochem.* 182:231-238, Hall *et al.* (1983) *J. Mol. Appl. Gen.* 2:101).

As used herein, growth under selective conditions means growth of a cell under conditions that require expression of a selectable marker for survival.

As used herein, an agent that destabilizes a chromosome is any  
20 agent known by those skilled in the art to enhance amplification events, and/or mutations. Such agents, which include BrdU, are well known to those skilled in the art.

In order to generate an artificial chromosome containing a particular heterologous nucleic acid of interest, it is possible to include the nucleic  
25 acid in the nucleic acid that is being introduced into cells to initiate production of the artificial chromosome. Thus, for example, a nucleic acid can be introduced into a cell along with nucleic acid encoding a selectable marker and/or a nucleic acid that targets to a heterochromatic region of a chromosome. For introducing a heterologous nucleic acid into

the cell, it can be included in a fragment that includes a selectable marker or as part of a separate nucleic acid fragment and introduced into the cell with a selectable marker during the process of generating the artificial chromosomes. Alternatively, heterologous nucleic acid can be introduced  
5 into an artificial chromosome at a later time after the initial generation of the artificial chromosome.

As used herein, the minichromosome refers to a chromosome derived from a multicentric, typically dicentric, chromosome that contains more euchromatic than heterochromatic DNA. For purposes herein, the  
10 minichromosome contains a *de novo* centromere (e.g., a neocentromere). In some embodiments, for example, the minichromosome contains a centromere that replicates in animals, e.g., a mammalian centromere or in plants, e.g., a plant centromere.

As used herein, *in vitro* assembled artificial chromosomes or  
15 synthetic chromosomes can be either more euchromatic than heterochromatic or more heterochromatic than euchromatic and are produced by joining essential components of a chromosome *in vitro*. These components include at least a centromere, a megareplicator, a telomere and optionally secondary origins of replication.

20 As used herein, *in vitro* assembled plant or animal artificial chromosomes are produced by joining essential components (at least the centromere, telomere(s), megareplicator and optional secondary origins of replication) that function in plants or animals. In particular embodiments, the megareplicator contains sequences of rDNA, particularly plant or  
25 animal rDNA.

As used herein, a plant is a eukaryotic organism that contains, in addition to a nucleus and mitochondria, chloroplasts capable of carrying out photosynthesis. A plant can be unicellular or multicellular and can contain multiple tissues and/or organs. Plants can reproduce sexually or

asexually and can be perennial or annual in growth. Plants can also be terrestrial or aquatic. The term "plant" includes a whole plant, plant cell, plant protoplast, plant calli, plant seed, plant organ, plant tissue, and other parts of a whole plant.

5           As used herein, stable maintenance of chromosomes occurs when at least about 85%, preferably 90%, more preferably 95%, of the cells retain the chromosome. Stability is measured in the presence of a selective agent. Preferably these chromosomes are also maintained in the absence of a selective agent. Stable chromosomes also retain their  
10 structure during cell culturing, suffering no unintended intrachromosomal or interchromosomal rearrangements.

          As used herein, *de novo* with reference to a centromere, refers to generation of an excess centromere in a chromosome as a result of incorporation of a heterologous nucleic acid fragment using the methods  
15 herein.

          As used herein, BrdU refers to 5-bromodeoxyuridine, which during replication is inserted in place of thymidine. BrdU is used as a mutagen; it also inhibits condensation of metaphase chromosomes during cell division.

20           As used herein, ribosomal RNA (rRNA) is the specialized RNA that forms part of the structure of a ribosome and participates in the synthesis of proteins. Ribosomal RNA is produced by transcription of genes which, in eukaryotic cells, are present in multiple copies. In human cells, the approximately 250 copies of rRNA genes (i.e., genes which encode rRNA)  
25 per haploid genome are spread out in clusters on at least five different chromosomes (chromosomes 13, 14, 15, 21 and 22). In mouse cells, the presence of ribosomal DNA (rDNA, which is DNA containing sequences that encode rRNA) has been verified on at least 11 pairs out of 20 mouse chromosomes (chromosomes 5, 6, 7, 9, 11, 12, 15, 16, 17, 18, and 19)

(see e.g., Rowe *et al.* (1996) *Mamm. Genome* 7:886-889 and Johnson *et al.* (1993) *Mamm. Genome* 4:49-52). In *Arabidopsis thaliana* the presence of rDNA has been verified on chromosomes 2 and 4 (18S, 5.8S, and 25S rDNA) and on chromosomes 3,4, and 5 (5S rDNA)(see The Arabidopsis Genome Initiative (2000) *Nature* 408:796-815). In eukaryotic cells, the multiple copies of the highly conserved rRNA genes are located in a tandemly arranged series of rDNA units, which are generally about 40-45 kb in length and contain a transcribed region and a nontranscribed region known as spacer (i.e., intergenic spacer) DNA which can vary in length and sequence. In the human and mouse, these tandem arrays of rDNA units are located adjacent to the pericentric satellite DNA sequences (heterochromatin). The regions of these chromosomes in which the rDNA is located are referred to as nucleolar organizing regions (NOR) which loop into the nucleolus, the site of ribosome production within the cell nucleus.

As used herein, a megachromosome refers to a chromosome that, except for introduced heterologous DNA, is substantially composed of heterochromatin. Megachromosomes are made up of an array of repeated amplicons that contain two inverted megareplicons bordered by introduced heterologous DNA (see, e.g., Figure 3 of U.S. Patent No. 6,077,697 for a schematic drawing of a megachromosome). For purposes herein, a megachromosome is about 50 to 400 Mb, generally about 250-400 Mb. Shorter variants are also referred to as truncated megachromosomes (about 90 to 120 or 150 Mb), dwarf megachromosomes (~150-200 Mb), and a micro-megachromosome (~50-90 Mb, typically 50-60 Mb). For purposes herein, the term

megachromosome refers to the overall repeated structure based on an array of repeated chromosomal segments (amplicons) that contain two inverted megareplicons bordered by any inserted heterologous DNA. The size will be specified.

5           As used herein, gene therapy involves the transfer or insertion of nucleic acid molecules into certain cells, which are also referred to as target cells, to produce specific products that are involved in preventing, curing, correcting, controlling or modulating diseases, disorders and deleterious conditions. The nucleic acid is introduced into the selected  
10 target cells in a manner such that the nucleic acid is expressed and a product encoded thereby is produced. Alternatively, the nucleic acid may in some manner mediate expression of DNA that encodes a therapeutic product. This product may be a therapeutic compound, which is produced in therapeutically effective amounts or at a therapeutically  
15 useful time. It may also encode a product, such as a peptide or RNA, that in some manner mediates, directly or indirectly, expression of a therapeutic product. Expression of the nucleic acid by the target cells within an organism afflicted with a disease or disorder thereby provides for modulation of the disease or disorder. The nucleic acid encoding the  
20 therapeutic product may be modified prior to introduction into the cells of the afflicted host in order to enhance or otherwise alter the product or expression thereof.

          For use in gene therapy, cells can be transfected *in vitro*, followed by introduction of the transfected cells into an organism. This is often  
25 referred to as *ex vivo* gene therapy. Alternatively, the cells can be transfected directly *in vivo* within an organism.

          As used herein, therapeutic agents include, but are not limited to, growth factors, antibodies, cytokines, such as tumor necrosis factors and interleukins, and cytotoxic agents and other agents disclosed herein and

known to those of skill in the art. Such agents include, but are not limited to, tumor necrosis factor,  $\alpha$ -interferon,  $\beta$ -interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 (IL-1), interleukin-2 (IL-2), interleukin-6 (IL-6), granulocyte  
5 macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), erythropoietin (EPO), pro-coagulants such as tissue factor and tissue factor variants, pro-apoptotic agents such FAS-ligand, fibroblast growth factors (FGF), nerve growth factor and other  
10 growth factors.

As used herein, a therapeutically effective product is a product that is encoded by heterologous DNA that, upon introduction of the DNA into a host, a product is expressed that effectively ameliorates or eliminates the symptoms, manifestations of an inherited or acquired disease or that  
15 cures the disease.

As used herein, transgenic plants and animals refer to plants and animals in which heterologous or foreign nucleic acid is expressed or in which the expression of a gene naturally present in the plant or animal has been altered by virtue of introduction of heterologous or foreign  
20 nucleic acid.

As used herein, IRES (internal ribosome entry site; see, *e.g.*, SEQ ID No. 27 and nucleotides 2736-3308 SEQ ID No. 28) refers to a region of a nucleic acid molecule, such as an mRNA molecule, that allows internal ribosome entry sufficient to initiate translation, which initiation  
25 can be detected in an assay for cap-independent translation (see, *e.g.*, U.S. Patent No. 6,171,821). The presence of an IRES within an mRNA molecule allows cap-independent translation of a linked protein-encoding sequence that otherwise would not be translated.

Internal ribosome entry site (IRES) elements were first identified in picornaviruses, which elements are considered the paradigm for cap-independent translation. The 5' UTRs of all picornaviruses are long and mediate translational initiation by directly recruiting and binding  
5 ribosomes, thereby circumventing the initial cap-binding step. IRES elements are frequently found in viral mRNA, they are rare in non-viral mRNA. Among non-viral mRNA molecules that contain functional IRES elements in their respective 5' UTRs are those encoding immunoglobulin heavy chain binding protein (BiP) (Macejak *et al.* (1991) *Nature*  
10 353:90-94); *Drosophila* Antennapedia (Oh *et al.* (1992) *Genes Dev.* 6:1643-1653); *D. Ultrabithorax* (Ye *et al.* (1997) *Mol. Cell Biol.* 17:1714-21); fibroblast growth factor 2 (Vagner *et al.* (1995) *Mol. Cell Biol.* 15:35-44); initiation factor eIF4G (Gan *et al.* (1998) *J. Biol. Chem.* 273:5006-5012); proto-oncogene c-myc (Nanbru *et al.* (1995) *J. Biol.*  
15 *Chem.* 272:32061-32066; Stoneley (1998) *Oncogene* 16:423-428); IRES<sub>H</sub>; from the 5'UTR of NRF1 gene (Oumard *et al.* (2000) *Mol. and Cell Biol.*, 20(8):2755-2759); and vascular endothelial growth factor (VEGF) (Stein *et al.* (1998) *Mol. Cell Biol.* 18:3112-9).

As used herein, a promoter, with respect to a region of DNA, refers  
20 to a sequence of DNA that contains a sequence of bases that signals RNA polymerase to associate with the DNA and initiate transcription of RNA (such as pol II for mRNA) from a template strand of the DNA. A promoter thus generally regulates transcription of DNA into mRNA. A particular promoter provided herein is the Ferritin heavy chain promoter (excluding  
25 the Iron Response Element, located in the 5'UTR), which was joined to the 37bp Fer-1 enhancer element. This promoter is set forth as SEQ ID NO:128. The endogenous Fer-1 enhancer element is located upstream of the Fer-1 promoter (e.g., a Fer-1 oligo was cloned proximal to the core promoter).

As used herein, isolated, substantially pure nucleic acid, such as, for example, DNA, refers to nucleic acid fragments purified according to standard techniques employed by those skilled in the art, such as that found in Sambrook *et al.* ((2001) Molecular Cloning: A Laboratory  
5 Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 3rd edition).

As used herein, expression refers to the transcription and/or translation of nucleic acid. For example, expression can be the transcription of a gene that may be transcribed into an RNA molecule,  
10 such as a messenger RNA (mRNA) molecule. Expression may further include translation of an RNA molecule and translated into peptides, polypeptides, or proteins. If the nucleic acid is derived from genomic DNA, expression may, if an appropriate eukaryotic host cell or organism is selected, include splicing of the mRNA. With respect to an antisense  
15 construct, expression may refer to the transcription of the antisense DNA.

As used herein, vector or plasmid refers to discrete elements that are used to introduce heterologous nucleic acids into cells for either expression of the heterologous nucleic acid or for replication of the heterologous nucleic acid. Selection and use of such vectors and  
20 plasmids are well within the level of skill of the art.

As used herein, transformation/transfection refers to the process by which nucleic acid is introduced into cells. The terms transfection and transformation refer to the taking up of exogenous nucleic acid, e.g., an expression vector, by a host cell whether or not any coding sequences  
25 are in fact expressed. Numerous methods of transfection are known to the ordinarily skilled artisan, for example, by *Agrobacterium*-mediated transformation, protoplast transformation (including polyethylene glycol (PEG)-mediated transformation, electroporation, protoplast fusion, and microcell fusion), lipid-mediated delivery, liposomes, electroporation,



sonoporation, microinjection, particle bombardment and silicon carbide whisker-mediated transformation and combinations thereof (see, *e.g.*, Paszkowski *et al.* (1984) *EMBO J.* 3:2717-2722; Potrykus *et al.* (1985) *Mol. Gen. Genet.* 199:169-177; Reich *et al.* (1986) *Biotechnology* 4:1001-1004; Klein *et al.* (1987) *Nature* 327:70-73; U.S. Patent No. 6,143,949; Paszkowski *et al.* (1989) in *Cell Culture and Somatic Cell Genetics of Plants, Vol. 6*, Molecular Biology of Plant Nuclear Genes, eds. Schell, J and Vasil, L.K. Academic Publishers, San Diego, California, p. 52-68; and Frame *et al.* (1994) *Plant J.* 6:941-948), direct uptake using calcium phosphate (CaPO<sub>4</sub>; see, *e.g.*, Wigler *et al.* (1979) *Proc. Natl. Acad. Sci. U.S.A.* 76:1373-1376), polyethylene glycol (PEG)-mediated DNA uptake, lipofection (see, *e.g.*, Strauss (1996) *Meth. Mol. Biol.* 54:307-327), microcell fusion (see, EXAMPLES, see, also Lambert (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:5907-5911; U.S. Patent No. 5,396,767, Sawford *et al.* (1987) *Somatic Cell Mol. Genet.* 13:279-284; Dhar *et al.* (1984) *Somatic Cell Mol. Genet.* 10:547-559; and McNeill-Killary *et al.* (1995) *Meth. Enzymol.* 254:133-152), lipid-mediated carrier systems (see, *e.g.*, Teifel *et al.* (1995) *Biotechniques* 19:79-80; Albrecht *et al.* (1996) *Ann. Hematol.* 72:73-79; Holmen *et al.* (1995) *In Vitro Cell Dev. Biol. Anim.* 31:347-351; Remy *et al.* (1994) *Bioconjug. Chem.* 5:647-654; Le Bolch *et al.* (1995) *Tetrahedron Lett.* 36:6681-6684; Loeffler *et al.* (1993) *Meth. Enzymol.* 217:599-618) or other suitable method. Methods for delivery of ACes are described in copending U.S. application Serial No. 09/815,979. Successful transfection is generally recognized by detection of the presence of the heterologous nucleic acid within the transfected cell, such as, for example, any visualization of the heterologous nucleic acid or any indication of the operation of a vector within the host cell.

As used herein, "delivery," which is used interchangeably with "transfection," refers to the process by which exogenous nucleic acid molecules are transferred into a cell such that they are located inside the cell. Delivery of nucleic acids is a distinct process from expression of  
5 nucleic acids.

As used herein, injected refers to the microinjection, such as by use of a small syringe, needle, or pipette, for injection of nucleic acid into a cell.

As used herein, substantially homologous DNA refers to DNA that  
10 includes a sequence of nucleotides that is sufficiently similar to another such sequence to form stable hybrids, with each other or a reference sequence, under specified conditions.

It is well known to those of skill in this art that nucleic acid fragments with different sequences may, under the same conditions,  
15 hybridize detectably to the same "target" nucleic acid. Two nucleic acid fragments hybridize detectably, under stringent conditions over a sufficiently long hybridization period, because one fragment contains a segment of at least about 10, 14 or 16 or more nucleotides in a sequence that is complementary (or nearly complementary) to a substantially  
20 contiguous sequence of at least one segment in the other nucleic acid fragment. If the time during which hybridization is allowed to occur is held constant, at a value during which, under preselected stringency conditions, two nucleic acid fragments with complementary base-pairing segments hybridize detectably to each other, departures from exact  
25 complementarity can be introduced into the base-pairing segments, and base-pairing will nonetheless occur to an extent sufficient to make hybridization detectable. As the departure from complementarity between the base-pairing segments of two nucleic acids becomes larger, and as

conditions of the hybridization become more stringent, the probability decreases that the two segments will hybridize detectably to each other.

Two single-stranded nucleic acid segments have "substantially the same sequence", if (a) both form a base-paired duplex with the same  
5 segment, and (b) the melting temperatures of the two duplexes in a solution of 0.5 X SSPE differ by less than 10°C. If the segments being compared have the same number of bases, then to have "substantially the same sequence", they will typically differ in their sequences at fewer than 1 base in 10. Methods for determining melting temperatures of  
10 nucleic acid duplexes are well known (see, *e.g.*, Meinkoth *et al.* (1984) *Anal. Biochem.* 138:267-284 and references cited therein).

As used herein, a nucleic acid probe is a DNA or RNA fragment that includes a sufficient number of nucleotides to specifically hybridize to DNA or RNA that includes complementary or substantially complementary  
15 sequences of nucleotides. A probe may contain any number of nucleotides, from as few as about 10 and as many as hundreds of thousands of nucleotides. The conditions and protocols for such hybridization reactions are well known to those of skill in the art as are the effects of probe size, temperature, degree of mismatch, salt  
20 concentration and other parameters on the hybridization reaction. For example, the lower the temperature and higher the salt concentration at which the hybridization reaction is carried out, the greater the degree of mismatch that may be present in the hybrid molecules.

To be used as a hybridization probe, the nucleic acid is generally  
25 rendered detectable by labeling it with a detectable moiety or label, such as  $^{32}\text{P}$ ,  $^3\text{H}$  and  $^{14}\text{C}$ , or by other means, including chemical labeling, such as by nick-translation in the presence of deoxyuridylate biotinylated at the 5'-position of the uracil moiety. The resulting probe includes the biotinylated uridylate in place of thymidylate residues and can be detected

(via the biotin moieties) by any of a number of commercially available detection systems based on binding of streptavidin to the biotin. Such commercially available detection systems can be obtained, for example, from Enzo Biochemicals, Inc. (New York, NY). Any other label known to those of skill in the art, including non-radioactive labels, may be used as long as it renders the probes sufficiently detectable, which is a function of the sensitivity of the assay, the time available (for culturing cells, extracting DNA, and hybridization assays), the quantity of DNA or RNA available as a source of the probe, the particular label and the means used to detect the label.

Once sequences with a sufficiently high degree of homology to the probe are identified, they can readily be isolated by standard techniques (see, *e.g.*, Sambrook *et al.* (2001) *Molecular Cloning: A Laboratory Manual*, 3rd Edition, Cold Spring Harbor Laboratory Press).

As used herein, conditions under which DNA molecules form stable hybrids are considered substantially homologous, and a DNA or nucleic acid homolog refers to a nucleic acid that includes a preselected conserved nucleotide sequence, such as a sequence encoding a polypeptide. By the term "substantially homologous" is meant having at least 75%, preferably 80%, preferably at least 90%, most preferably at least 95% homology therewith or a less percentage of homology or identity and conserved biological activity or function.

The terms "homology" and "identity" are often used interchangeably. In this regard, percent homology or identity may be determined, for example, by comparing sequence information using a GAP computer program. The GAP program utilizes the alignment method of Needleman and Wunsch (*J. Mol. Biol.* 48:443 (1970), as revised by Smith and Waterman (*Adv. Appl. Math.* 2:482 (1981)). Briefly, the GAP program defines similarity as the number of aligned symbols (i.e.,

nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences. The preferred default parameters for the GAP program may include: (1) a unary comparison matrix (containing a value of 1 for identities and 0 for non-identities) and  
5 the weighted comparison matrix of Gribskov and Burgess, *Nucl. Acids Res.* 14:6745 (1986), as described by Schwartz and Dayhoff, eds., *ATLAS OF PROTEIN SEQUENCE AND STRUCTURE*, National Biomedical Research Foundation, pp. 353-358 (1979); (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3)  
10 no penalty for end gaps.

By sequence identity, the number of conserved amino acids are determined by standard alignment algorithms programs, and are used with default gap penalties established by each supplier. Substantially homologous nucleic acid molecules would hybridize typically at moderate  
15 stringency or at high stringency all along the length of the nucleic acid of interest. Preferably the two molecules will hybridize under conditions of high stringency. Also contemplated are nucleic acid molecules that contain degenerate codons in place of codons in the hybridizing nucleic acid molecule.

20 Whether any two nucleic acid molecules have nucleotide sequences that are at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% "identical" can be determined using known computer algorithms such as the "FAST A" program, using for example, the default parameters as in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 (1988).  
25 Alternatively the BLAST function of the National Center for Biotechnology Information database may be used to determine relative sequence identity.

In general, sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be

calculated using published techniques. (See, e.g.: *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H. & Lipton, D., *SIAM J Applied Math* 48:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H. & Lipton, D., *SIAM J Applied Math* 48:1073 (1988). Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(II):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., et al., *J Molec Biol* 215:403 (1990)).

Therefore, as used herein, the term "identity" represents a comparison between a test and a reference polypeptide or polynucleotide. For example, a test polypeptide may be defined as any polypeptide that is 90% or more identical to a reference polypeptide.

As used herein, the term at least "90% identical to" refers to percent identities from 90 to 99.99 relative to the reference polypeptides. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of

100 amino acids are compared. No more than 10% (i.e., 10 out of 100) amino acids in the test polypeptide differs from that of the reference polypeptides. Similar comparisons may be made between a test and reference polynucleotides. Such differences may be represented as point mutations randomly distributed over the entire length of an amino acid sequence or they may be clustered in one or more locations of varying length up to the maximum allowable, e.g. 10/100 amino acid difference (approximately 90% identity). Differences are defined as nucleic acid or amino acid substitutions, or deletions.

10 As used herein: stringency of hybridization in determining percentage mismatch encompass the following conditions or equivalent conditions thereto:

- 1) high stringency: 0.1 x SSPE or SSC, 0.1% SDS, 65°C
- 2) medium stringency: 0.2 x SSPE or SSC, 0.1% SDS, 50°C
- 15 3) low stringency: 1.0 x SSPE or SSC, 0.1% SDS, 50°C

or any combination of salt and temperature and other reagents that result in selection of the same degree of mismatch or matching. Equivalent conditions refer to conditions that select for substantially the same percentage of mismatch in the resulting hybrids. Additions of ingredients, such as formamide, Ficoll, and Denhardt's solution affect parameters such as the temperature under which the hybridization should be conducted and the rate of the reaction. Thus, hybridization in 5 X SSC, in 20% formamide at 42° C is substantially the same as the conditions recited above hybridization under conditions of low stringency. The recipes for SSPE, SSC and Denhardt's and the preparation of deionized formamide are described, for example, in Sambrook *et al.* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Chapter 8; see, Sambrook *et al.*, vol. 3, p. B.13, see, also, numerous catalogs that describe commonly used laboratory solutions. It is understood that

equivalent stringencies may be achieved using alternative buffers, salts and temperatures. As used herein, all assays and procedures, such as hybridization reactions and antibody-antigen reactions, unless otherwise specified, are conducted under conditions recognized by those of skill in the art as standard conditions.

As used herein, conservative amino acid substitutions, such as those set forth in Table 1, are those that do not eliminate biological activity. Suitable conservative substitutions of amino acids are known to those of skill in this art and may be made generally without altering the biological activity of the resulting molecule. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (see, e.g., Watson *et al. Molecular Biology of the Gene*, 4th Edition, 1987, The Bejacmin/Cummings Pub. co., p.224). Conservative amino acid substitutions are made, for example, in accordance with those set forth in TABLE 1 as follows:

TABLE 1

	Original residue	Conservative substitution
	Ala (A)	Gly; Ser, Abu
20	Arg (R)	Lys, orn
	Asn (N)	Gln; His
	Cys (C)	Ser
	Gln (Q)	Asn
	Glu (E)	Asp
25	Gly (G)	Ala; Pro
	His (H)	Asn; Gln
	Ile (I)	Leu; Val; Met; Nle; Nva
	Leu (L)	Ile; Val; Met; Nle; Nva
	Lys (K)	Arg; Gln; Glu
30	Met (M)	Leu; Tyr; Ile; NLe Val
	Ornithine	Lys; Arg
	Phe (F)	Met; Leu; Tyr
	Ser (S)	Thr
	Thr (T)	Ser
35	Trp (W)	Tyr
	Tyr (Y)	Trp; Phe
	Val (V)	Ile; Leu; Met; Nle; Nva



Other substitutions are also permissible and may be determined empirically or in accord with known conservative substitutions.

As used herein, the amino acids, which occur in the various amino acid sequences appearing herein, are identified according to their well-known, three-letter or one-letter abbreviations. The nucleotides, which occur in the various DNA fragments, are designated with the standard single-letter designations used routinely in the art.

As used herein, a splice variant refers to a variant produced by differential processing of a primary transcript of genomic DNA that results in more than one type of mRNA.

As used herein, a probe or primer based on a nucleotide sequence includes at least 10, 14, 16, 30 or 100 contiguous nucleotides from the reference nucleic acid molecule.

As used herein, recombinant production by using recombinant DNA methods refers to the use of the well known methods of molecular biology for expressing proteins encoded by cloned DNA.

As used herein, biological activity refers to the *in vivo* activities of a compound or physiological responses that result upon *in vivo* administration of a compound, composition or other mixture. Biological activity, thus, encompasses therapeutic effects and pharmaceutical activity of such compounds, compositions and mixtures. Biological activities may be observed in *in vitro* systems designed to test or use such activities. Thus, for purposes herein the biological activity of a luciferase is its oxygenase activity whereby, upon oxidation of a substrate, light is produced.

The terms substantially identical or similar varies with the context as understood by those skilled in the relevant art and generally means at least 40, 60, 80, 90, 95 or 98%.

As used herein, substantially identical to a product means  
5 sufficiently similar so that the property is sufficiently unchanged so that the substantially identical product can be used in place of the product.

As used herein, substantially pure means sufficiently homogeneous to appear free of readily detectable impurities as determined by standard methods of analysis, such as thin layer chromatography (TLC), gel  
10 electrophoresis and high performance liquid chromatography (HPLC), used by those of skill in the art to assess such purity, or sufficiently pure such that further purification would not detectably alter the physical and chemical properties, such as enzymatic and biological activities, of the substance. Methods for purification of the compounds to produce  
15 substantially chemically pure compounds are known to those of skill in the art. A substantially chemically pure compound may, however, be a mixture of stereoisomers or isomers. In such instances, further purification might increase the specific activity of the compound.

As used herein, vector (or plasmid) refers to discrete elements that  
20 are used to introduce heterologous DNA into cells for either expression or replication thereof. The vectors typically remain episomal, but may be designed to effect integration of a gene or portion thereof into a chromosome of the genome. Also contemplated are vectors that are artificial chromosomes, such as yeast artificial chromosomes and  
25 mammalian artificial chromosomes. Selection and use of such vehicles are well known to those of skill in the art. An expression vector includes vectors capable of expressing DNA that is operatively linked with regulatory sequences, such as promoter regions, that are capable of effecting expression of such DNA fragments. Thus, an expression vector

refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector that, upon introduction into an appropriate host cell, results in expression of the cloned DNA.

Appropriate expression vectors are well known to those of skill in the art and include those that are replicable in eukaryotic cells and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome.

As used herein, protein-binding-sequence refers to a protein or peptide sequence that is capable of specific binding to other protein or peptide sequences generally, to a set of protein or peptide sequences or to a particular protein or peptide sequence.

As used herein, a composition refers to any mixture of two or more ingredients. It may be a solution, a suspension, liquid, powder, a paste, aqueous, non-aqueous or any combination thereof.

As used herein, a combination refers to any association between two or more items.

As used herein, fluid refers to any composition that can flow. Fluids thus encompass compositions that are in the form of semi-solids, pastes, solutions, aqueous mixtures, gels, lotions, creams and other such compositions.

As used herein, a cellular extract refers to a preparation or fraction that is made from a lysed or disrupted cell.

As used herein, the term "subject" refers to animals, plants, insects, and birds and other phyla, genera and species into which nucleic acid molecules may be introduced. Included are higher organisms, such as mammals, fish, insects and birds, including humans, primates, cattle, pigs, rabbits, goats, sheep, mice, rats, guinea pigs, hamsters, cats, dogs, horses, chicken and others.

As used herein, flow cytometry refers to processes that use a laser based instrument capable of analyzing and sorting out cells and or chromosomes based on size and fluorescence.

As used herein, the abbreviations for any protective groups, amino  
5 acids and other compounds, are, unless indicated otherwise, in accord with their common usage, recognized abbreviations, or the IUPAC-IUB Commission on Biochemical Nomenclature (see, (1972) *Biochem.* 11:942-944).

**B. Recombination systems**

10 Site-specific recombination systems typically contain three elements: a pair of DNA sequences (the site-specific recombination sequences) and a specific enzyme (the site-specific recombinase). The site-specific recombinase catalyzes a recombination reaction between two site-specific recombination sequences.

15 A number of different site-specific recombinase systems are available and/or known to those of skill in the art, including, but not limited to: the Cre//lox recombination system using CRE recombinase (see, e.g., SEQ ID Nos. 58 and 59) from the Escherichia coli phage P1 (see, e.g., Sauer (1993) *Methods in Enzymology* 225:890-900; Sauer et al.  
20 (1990) *The New Biologist* 2:441-449), Sauer (1994) *Current Opinion in Biotechnology* 5:521-527; Odell et al. (1990) *Mol Gen Genet.* 223:369-378; Lasko et al. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:6232-6236; U.S. Patent No. 5,658,772), the FLP/FRT system of yeast using the FLP recombinase (see, SEQ ID Nos. 60 and 61) from the 2 $\mu$  episome of  
25 *Saccharomyces cerevisiae* (Cox (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80:4223; Falco et al. (1982) *Cell* 29:573-584; Golic et al. (1989) *Cell* 59:499-509; U.S. Patent No. 5,744,336), the resolvases, including Gin recombinase of phage Mu (Maeser et al. (1991) *Mol Gen Genet.* 230:170-176; Klippel, A. et al (1993) *EMBO J.* 12:1047-1057; see, e.g.,

SEQ ID Nos. 64-67), Cin, Hin,  $\alpha\delta$  Tn3; the Pin recombinase of *E. coli* (see, e.g., SEQ ID Nos. 68 and 69; Enomoto *et al.* (1983) *J. Bacteriol.* 6:663-668), the R/RS system of the pSR1 plasmid of *Zygosaccharomyces rouxii* (Araki *et al.* (1992) *J. Mol. Biol.* 225:25-37; Matsuzaki *et al.* (1990) *J. Bacteriol.* 172: 610-618) and site-specific recombinases from *Kluyveromyces drosophilarius* (Chen *et al.* (1986) *Nucleic Acids Res.* 314:4471-4481) and *Kluyveromyces waltii* (Chen *et al.* (1992) *J. Gen. Microbiol.* 138:337-345). Other systems are known to those of skill in the art (Stark *et al. Trends Genet.* 8:432-439; Utatsu *et al.* (1987) *J. Bacteriol.* 169:5537-5545; see, also, U.S. Patent No. 6,171,861).

Members of the highly related family of site-specific recombinases, the resolvase family, such as  $\gamma\delta$ , Tn3 resolvase, Hin, Gin, and Cin are also available. Members of this family of recombinases are typically constrained to intramolecular reactions (e.g., inversions and excisions) and can require host-encoded factors. Mutants have been isolated that relieve some of the requirements for host factors (Maeser *et al.* (1991) *Mol. Gen. Genet.* 230:170-176), as well as some of the constraints of intramolecular recombination (see, U.S. Patent No. 6,171,861).

The bacteriophage P1 Cre/lox and the yeast FLP/FRT systems are particularly useful systems for site-specific integration, inversion or excision of heterologous nucleic acid into, and out of, chromosomes, particularly ACes as provided herein. In these systems a recombinase (Cre or FLP) interacts specifically with its respective site-specific recombination sequence (lox or FRT, respectively) to invert or excise the intervening sequences. The sequence for each of these two systems is relatively short (34 bp for lox and 47 bp for FRT).

The FLP/FRT recombinase system has been demonstrated to function efficiently in plant cells (U.S. Patent No. 5,744,386), and, thus, can be used for producing plant artificial chromosome platforms. In

general, short incomplete FRT sites leads to higher accumulation of excision products than the complete full-length FRT sites. The system catalyzes intra- and intermolecular reactions, and, thus, can be used for DNA excision and integration reactions. The recombination reaction is  
5 reversible and this reversibility can compromise the efficiency of the reaction in each direction. Altering the structure of the site-specific recombination sequences is one approach to remedying this situation. The site-specific recombination sequence can be mutated in a manner that the product of the recombination reaction is no longer recognized as  
10 a substrate for the reverse reaction, thereby stabilizing the integration or excision event.

In the Cre-lox system, discovered in bacteriophage P1, recombination between loxP sites occurs in the presence of the Cre recombinase (see, *e.g.*, U.S. Patent No. 5,658,772). This system can be  
15 used to insert, invert or excise nucleic acid located between two lox sites. Cre can be expressed from a vector. Since the lox site is an asymmetrical nucleotide sequence, lox sites on the same DNA molecule can have the same or opposite orientation with respect to each other. Recombination between lox sites in the same orientation results in a deletion of the DNA  
20 segment located between the two lox sites and a connection between the resulting ends of the original DNA molecule. The deleted DNA segment forms a circular molecule of DNA. The original DNA molecule and the resulting circular molecule each contain a single lox site. Recombination between lox sites in opposite orientations on the same DNA molecule  
25 result in an inversion of the nucleotide sequence of the DNA segment located between the two lox sites. In addition, reciprocal exchange of DNA segments proximate to lox sites located on two different DNA molecules can occur. All of these recombination events are catalyzed by the product of the Cre coding region.

Any site-specific recombinase system known to those of skill in the art is contemplated for use herein. It is contemplated that one or a plurality of sites that direct the recombination by the recombinase are introduced into an artificial chromosome to produce platform ACes. The resulting platform ACes are introduced into cells with nucleic acid encoding the cognate recombinase, typically on a vector, and nucleic acid encoding heterologous nucleic acid of interest linked to the appropriate recombination site for insertion into the platform ACes. The recombinase-encoding-nucleic acid may be introduced into the cells on the same vector, or a different vector, encoding the heterologous nucleic acid.

An *E. coli* phage lambda integrase system for ACes platform engineering and for artificial chromosome engineering is provided (Lorbach *et al.* (2000) *J. Mol. Biol.* 296:1175-1181). The phage lambda integrase (Landy, A. (1989) *Annu. Rev. Biochem.* 58:913-94) is adapted herein and the cognate *att* sites are provided. Chromosomes, including ACes, engineered to contain one or a plurality of *att* sites are provided, as are vectors encoding a mutant integrase that functions in the absence other factors. Methods using the modified chromosomes and vectors for introduction of heterologous nucleic acid are also provided.

For purposes herein, one or more of the sites (e.g., a single site or a pair of sites) required for recombination are introduced into an artificial chromosome, such as an ACes chromosome. The enzyme for catalyzing site-directed recombination is introduced with the DNA of interest, or separately, or is engineered onto the artificial chromosome under the control of a regulatable promoter.

As described herein, artificial chromosome platforms containing one or multiple recombination sites are provided. The methods and resulting products are exemplified with the lambda phage Att/Int system, but

similar methods may be used for production of *ACes* platforms with other recombination systems.

The Att/Int system and vectors provided herein are not only intended for engineering *ACes* platforms, but may be used to engineer an Att/Int system into any chromosome. Introduction of att sites into a chromosome will permit engineering of natural chromosomes, such as by permitting targeted integration genes or regulatory regions, and by controlled excision of selected regions. For example, genes encoding a particular trait may be added to a chromosome, such as plant chromosome engineered to contain one or plurality of *att* sites. Such chromosomes may be used for screening DNA to identify genes. Large pieces of DNA can be introduced into cells and the cells screened phenotypically to select those having the desired trait.

#### **C. Platforms**

Provided herein are platform artificial chromosomes (platform *ACes*) containing single or multiple site-specific recombination sites. Chromosome-based platform technology permits efficient and tractable engineering and subsequent expression of multiple gene targets. Methods are provided that use DNA vectors and fragments to create platform artificial chromosomes, including animal, particularly mammalian, artificial chromosomes, and plant artificial chromosomes. The artificial chromosomes contain either single or multiple sequence-specific recombination sites suitable for the placement of target gene expression vectors onto the platform chromosome. The engineered chromosome-based platform *ACes* technology is applicable for methods, including cellular and transgenic protein production, transgenic plant and animal production and gene therapy. The platform *ACes* are also useful for producing a library of *ACes* comprising random portions of a given genome (e.g., a mammalian, plant or prokaryotic genome) for genomic



screening; as well as a library of cells comprising different and/or mutually exclusive *ACes* therein.

Exemplary of artificial chromosome platforms are those based on *ACes*. *ACes* artificial chromosomes are non-viral, self-replicating nucleic acid molecules that function as a natural chromosome, having all the elements required for normal chromosomal replication and maintenance within the cell nucleus. *ACes* artificial chromosomes do not rely on integration into the genome of the cell to be effective, and they are not limited by DNA carrying capacity and as such the therapeutic gene(s) of interest, including regulatory sequences, can be engineered into the *ACes*. In addition, *ACes* are stable *in vitro* and *in vivo* and can provide predictable long-term gene expression. Once engineered and delivered to the appropriate cell or embryo, *ACes* work independently alongside host chromosomes, for *ACes* that are predominantly heterochromatin producing only the products (proteins) from the genes it carries. As provided herein *ACes* are modified by introduction of recombination site(s) to provide a platform for ready introduction of heterologous nucleic acid. The *ACes* platforms can be used for production of transgenic animals and plants; as vectors for genetic therapy; for use as protein production systems; for animal models to identify and target new therapeutics; in cell culture for the development and production of therapeutic proteins; and for a variety of other applications.

#### **1. Generation of artificial chromosomes**

Artificial chromosomes may be generated by any method known to those of skill in the art. Of particular interest herein are the *ACes* artificial chromosomes, which contain a repeated unit. Methods for production of *ACes* are described in detail in U.S. Patent Nos. 6,025,155 and 6,077,697, which, as with all patents, applications, publications and other disclosure, are incorporated herein in their entirety.

**Generation of *de novo* ACes.**

ACes can be generated by cotransfecting exogenous DNA—such as a mammary tissue-specific DNA cassette including the gene sequences for a therapeutic protein, with a rDNA fragment and a drug resistance marker gene into the desired eukaryotic cell, such as plant or animal cells, such as murine cells *in vitro*. DNA with a selectable or detectable marker is introduced, and can be allowed to integrate randomly into pericentric heterochromatin or can be targeted to pericentric heterochromatin, such as that in rDNA gene arrays that reside on acrocentric chromosomes, such as the short arms of acrocentric chromosomes. This integration event activates the “megareplicator” sequence and amplifies the pericentric heterochromatin and the exogenous DNA, and duplicates a centromere. Ensuing breakage of this “dicentric” chromosome can result in the production of daughter cells that contain the substantially-original chromosome and the new artificial chromosome. The resulting ACes contain all the essential elements needed for stability and replication in dividing cells—centromere, origins of replications, and telomeres. ACes have been produced that express marker genes (lacZ, green fluorescent protein, neomycin-resistance, puromycin-resistance, hygromycin-resistance) and genes of interest. Isolated ACes, for example, have been successfully transferred intact to rodent, human, and bovine cells by electroporation, sonoporation, microinjection, and transfection with lipids and dendrimers.

To render the creation of ACes with desired genes more tractable and efficient, “platform” ACes (platform-ACes) can be produced that contain defined DNA sequences for enzyme-mediated homologous DNA recombination, such as by Cre or FLP recombinases (Bouhassira *et al.* (1996) *Blood* 88(supplement 1):190a; Bouhassira *et al.* (1997) *Blood*, 90:3332-3344; Siebler *et al.* (1997) *Biochemistry*: 36:1740-1747;

Siebler *et al.* (1998) *Biochemistry* 37: 6229-6234; and Bethke *et al.* (1997) *Nucl. Acids Res.* 25:2828-2834), and as exemplified herein the lambda phage integrase. A *lox* site contains two 13 bp inverted repeats to which Cre-recombinase binds and an intervening 8 bp core region.

- 5 Only pairs of sites having identity in the central 6 bp of the core region are proficient for recombination; sites having non-identical core sequences (heterospecific *lox* sites) do not efficiently recombine with each other (Hoess *et al.* (1986) *Nucleic Acids Res.* 14:2287-2300).

10 **Generating acrocentric chromosomes for plant artificial chromosome formation.**

- In human and mouse cells *de novo* formation of a satellite DNA based artificial chromosome (SATAC, also referred to as *ACes*) can occur in an acrocentric chromosome where the short arm contains only pericentric heterochromatin, the rDNA array, and telomere sequences.
- 15 Plant species may not have any acrocentric chromosomes with the same physical structure described, but "megareplicator" DNA sequences reside in the plant rDNA arrays, also known as the nucleolar organizing regions (NOR). A structure like those seen in acrocentric mammalian chromosomes can be generated using site-specific recombination between
- 20 appropriate arms of plant chromosomes.

**Approach**

- Qin *et al.* ((1994) *Proc. Natl. Acad. Sci. U.S.A.* 91:1706-1710, 1994) describes crossing two *Nicotiana tabacum* transgenic plants. One plant contains a construct encoding a promoterless hygromycin-resistance
- 25 gene preceded by a *lox* site (*lox-hpt*), the other plant carries a construct containing a cauliflower mosaic virus 35S promoter linked to a *lox* sequence and the *cre* DNA recombinase coding region (35S-*lox-cre*). The constructs were introduced separately by infecting leaf explants with *agrobacterium tumefaciens* which carries the kanamycin-resistance gene

(Kan<sup>R</sup>). The resultant Kan<sup>R</sup> transgenic plants were crossed. Plants that carried the appropriate DNA recombination event were identified by hygromycin-resistance.

5                   **Modification of the above for generation of ACes**

The Kan<sup>R</sup> cultivars are initially screened, such as by FISH, to identify two sets of candidate transgenic plants. One set has one construct integrated in regions adjacent to the pericentric heterochromatin on the short arm of any chromosome. The second set of candidate plants  
10 has the other construct integrated in the NOR region of appropriate chromosomes. To obtain reciprocal translocation both sites must be in the same orientation. Therefore a series of crosses are required, Kan<sup>R</sup> plants generated, and FISH analyses performed to identify the appropriate "acrocentric" plant chromosome for *de novo* plant ACes formation.

15           2.       **Bacteriophage lambda integrase-based site-specific recombination system**

An integral part of the platform technology includes a site-specific recombination system that allows the placement of selected gene targets or genomic fragments onto the platform chromosomes. Any such system  
20 may be used. In particular, a method is provided for insertion of additional DNA fragments into the platform chromosome residing in the cell via sequence-specific recombination using the recombinase activity of the bacteriophage lambda integrase. The lambda integrase system is exemplary of the recombination systems contemplated for ACes. Any  
25 known recombination system, including any described herein, particularly any that operates without the need for additional factors or that, by virtue of mutation, does not require additional factors, is contemplated.

As noted the lambda integrase system provided herein can be used with natural chromosomes and artificial chromosomes in addition to ACes. Single or a plurality of recombination sites, which may be the same or different, are introduced into artificial chromosomes to produce  
5 artificial chromosome platforms.

**3. Creation of bacteriophage lambda integrase site-specific recombination system**

The lambda phage-encoded integrase (designated Int) is a prototypical member of the integrase family. Int effects integration and  
10 excision of the phage in and out of the *E. coli* genome via recombination between pairs of attachment sites designated *attB/attP* and *attL/attR*. Each *att* site contains two inverted 9 base pair core Int binding sites and a 7 base pair overlap region that is identical in wild-type *att* sites. Each  
15 site, except for *attB* contains additional Int binding sites. In flanking regions, there are recognition sequences for accessory DNA binding proteins, such as integration host factor (IHF), factor for inversion stimulation (FIS) and the phage encoded excision protein (XIS). Except  
20 for *attB*, Int is a heterobivalent DNA-binding protein and, with assistance from the accessory proteins and negative DNA supercoiling, binds simultaneously to core and arm sites within the same *att* site.

Int, like Cre and FLP, executes an ordered sequential pair of strand exchanges during integrative and excisive recombination. The natural pairs of target sequences for Int, *attB* and *attP* or *attL* and *attR* are located on the same or different DNA molecules resulting in intra or  
25 intermolecular recombination, respectively. For example, intramolecular recombination occurs between inversely oriented *attB* and *attP*, or between *attL* and *attR* sequences, respectively, leading to inversion of the intervening DNA segment.

Like the recombinase systems, such as Cre and FLP, Int directs site-specific recombination. Unlike the other systems, such as Cre and FLP, Int generally requires additional protein factors for integrative and excisive recombination and negative supercoiling for integrative recombination.

5 Hence, the Int system had not been used in eukaryotic targeting systems.

Mutant Int proteins, designated Int-h (E174K) and a derivative thereof Int-h/218(E174K/E218K) do not require accessory proteins to perform intramolecular integrative and excisive recombination in co-transfection assays in human cells (Lorbach *et al.* (2000) *J Mol. Biol.*  
10 296:1175-1181); wild-type Int does not catalyze intramolecular recombination in human cells harboring target sites *attB* and *attP*. Hence it had been demonstrated that mutant Int can catalyze factor-independent recombination events in human cells.

There has been no demonstration by others that this system can be  
15 used for engineering of eukaryotic genomes or chromosomes. Provided herein are chromosomes, including artificial chromosomes, such as but not limited to *ACes* that contain *att* sites (e.g., platform *ACes*), and the use of such chromosomes for targeted integration of heterologous DNA into such chromosomes in eukaryotic cells, including animal, such as  
20 rodent and human, and plant cells. Mutant Int provided herein is shown to effect site-directed recombination between sites in artificial chromosomes and vectors containing cognate sites.

An additional component of the chromosome-based platform technology is the site-specific integration of target DNA sequences onto  
25 the platform. For this the native bacteriophage lambda integrase has been modified to carry out this sequence specific DNA recombination event in eukaryotic cells. The bacteriophage lambda integrase and its cognate DNA substrate *att* is a member of the site-specific recombinase family that also includes the bacteriophage P1 Cre/lox system as well as

the *Saccharomyces cerevisiae* 2 micron based FLP/FRT system (see, e.g., Landy (1989) *Ann. Rev. Biochem* 58:913-949; Hoess *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:3398-3402; Broach *et al.* (1982) *Cell* 29:227-234).

5 By combining DNA endonuclease and DNA ligase activity these recombinases recognize and catalyze DNA exchanges between sequences flanking the recognition site. During the integration of lambda genome into the *E. coli* (lambda recombination) genome, the phage integrase (INT) in association with accessory proteins catalyzes the DNA exchange  
10 between the attP site of the phage genome and the attB site of the bacterial genome resulting in the formation of attL and attR sites (Figure 6). The engineered bacteriophage lambda integrase has been produced herein to carry out an intermolecular DNA recombination event between  
15 an incoming DNA molecule (primarily on a vector containing the bacterial attB site) and the chromosome-based platform carrying the lambda attP sequence independent of lambda bacteriophage or bacterial accessory proteins.

In contrast to the bi-directional Cre/lox and FLP/FRT system, the engineered lambda recombination system derived for chromosome-based  
20 platform technology is advantageously unidirectional because accessory proteins, which are absent, are required for excision of integrated nucleic acid upon further exposure to the lambda Int recombinase.

**4. Creation of platform chromosome containing single or multiple sequence-specific recombination sites**

25 **a. Multiple sites**

For the creation of a platform chromosome containing multiple, sequence-specific recombination sites, artificial chromosomes are produced as depicted in Figure 5 and Example 3. As discussed above, artificial chromosomes can be produced using any suitable methodology,

including those described in U.S. Patent Nos. 5,288,625; 5,712,134; 5,891,691; 6,025,155. Briefly, to prepare artificial chromosomes containing multiple recombination (e.g., integration) sites, nucleic acid (either in the form a one or more plasmids, such as the plasmid pSV40193attPsensePUR set forth in Example 3) is targeted into an amplifiable region of a chromosome, such as the pericentric region of a chromosome. Among such regions are the rDNA gene loci in acrocentric mammalian chromosomes. Hence, targeting nucleic acid for integration into the rDNA region of mammalian acrocentric chromosomes can include the mouse rDNA fragments (for targeting into rodent cell lines) or large human rDNA regions on BAC/PAC vectors (or subclones thereof in standard vectors) for targeting into human acrocentric chromosomes, such as for human gene therapy applications. The targeting nucleic acid generally includes a detectable or selectable marker, such as antibiotic resistance, such as puromycin and hygromycin, a recombination site (such as attP, attB, attL, attR or the like), and/or human selectable markers as required for gene therapy applications. Cells are grown under conditions that result in amplification and ultimately production of ACes artificial chromosomes having multiple recombination (e.g., integration) sites therein. ACes having the desired size are selected for further engineering.

**b. Creation of platform chromosome containing a single sequence-specific recombination site**

In this method a mammalian platform artificial chromosome is generated containing a single sequence-specific recombination site. In the Example below, this approach is demonstrated using a puromycin resistance marker for selection and a mouse rDNA fragment for targeting into the rDNA locus on mouse acrocentric chromosomes. Other selection markers and targeting DNA sequences as desired and known to those of



skill in the art can be used. Additional resistance markers include genes conferring resistance to the antibiotics neomycin, blasticidin, hygromycin and zeocin. For applications, such as gene therapy in which potentially immunogenic responses are to be avoided, host, such as human, derived

5 selectable markers or markers detectable with monoclonal antibodies (MAb) followed by fluorescent activated cell sorting (FACS) can be used. Examples in this class include, but are not limited to: human nerve growth factor receptor (detection with MAb); truncated human growth factor receptor (detection with MAb); mutant human dihydrofolate reductase

10 (DHFR; detectable using a fluorescent methotrexate substrate); secreted alkaline phosphatase (SEAP; detectable with fluorescent substrate); thymidylate synthase (TS; confers resistance to fluorodeoxyuridine); human CAD gene (confers resistance to N-phosphonacetyl-L-aspartate (PALA)).

15 To construct a platform artificial chromosome with a single site, an ACes artificial chromosome (or other artificial chromosome of interest) can be produced containing a selectable marker. A single sequence specific recombination site is targeted onto ACes via homologous recombination. For this, DNA sequences containing the site-specific

20 recombination sequence are flanked with DNA sequences homologous to a selected sequence in the chromosome. For example, when using a chromosome containing rDNA or satellite DNA, such DNA can be used as homologous sequences to target the site-specific recombination sequence onto the chromosome. A vector is designed to have these homologous

25 sequences flanking the site-specific recombination site and, after the appropriate restriction enzyme digest to generate free ends of homology to the chromosome, the DNA is transfected into cells harboring the chromosome. After transfection and integration of the site-specific cassette, homologous recombination events onto the platform

chromosome are subcloned and identified, for example by screening single cell subclones via expression of resistance or a fluorescent marker and PCR analysis. In one embodiment, a platform artificial chromosome, such as a platform ACes, that contains a single copy of the recombination site is selected. Examples 2B and 2D exemplify the process, and Figure 3 provides a diagram depicting one method for the creation of a platform mammalian chromosome containing a single sequence-specific recombination site.

10        **5.        Lambda integrase mediated recombination of target gene expression vector onto platform chromosome**

          The third component of the chromosome-based platform technology involves the use of target gene expression vectors carrying, for example, genes for gene therapy, genes for transgenic animal or plant production, and those required for cellular protein production of interest.

15    Using lambda integrase mediated site-specific recombination, or any other recombinase-mediated site-specific recombination, the target gene expression vectors are introduced onto the selected chromosome platform. The use of target gene expression vector permits use of the *de novo* generated chromosome-based platforms for a wide range of gene

20    targets. Furthermore, chromosome platforms containing multiple *attP* sites provides the opportunity to incorporate multiple gene targets onto a single platform, thereby providing for expression of multiple gene targets, including the expression of cellular and genetic regulatory genes and the expression of all or parts of metabolic pathways. In addition to

25    expressing small target genes, such as cDNA and hybrid cDNA/artificial intron constructs, the chromosome-based platform can be used for engineering and expressing large genomic fragments carrying target genes along with its endogenous genomic promoter sequences. This is of importance, for example, where the therapy requires precise cell specific

expression and in instances where expression is best achieved from genomic clones rather than cDNA clones. Figure 9 provides a diagram summarizing one embodiment of the chromosome-based technology.

A feature of the target gene expression vector that is of interest to  
5 include is a promoterless marker gene, which as exemplified (see, Figure 9) contains an upstream *attB* site (marker 2 on Figure 9). The nucleic acid encoding the marker is not expressed unless it is placed downstream from a promoter sequence. Using the recombinase technology provided herein, such as the lambda integrase technology ( $\lambda$ INT<sub>E174R</sub> on figure 8)  
10 provided herein, site-specific recombination between the *attB* site on the vector and the promoter-*attP* site (in the "sense" orientation) on the chromosome-based platform results in the expression of marker 2 on the target gene expression vector, thereby providing a positive selection for the lambda INT mediated site-specific recombination event. Site-specific  
15 recombination events on the chromosome-based platform versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR. Examples of suitable marker 2 genes, include, but are not limited to, genes that confer resistance to toxic compounds or antibiotics,  
20 fluorescence activated cell sorting (FACS) sortable cell surface markers and various fluorescent markers. Examples of these genes include, but are not limited to, human L26a<sup>R</sup> (human homolog of *Saccharomyces cerevisiae* CYH<sup>B</sup> gene), neomycin, puromycin, blasticidin, CD24 (see, e.g., US Patents 5,804,177 and 6,074,836), truncated CD4, truncated low  
25 affinity nerve growth factor receptor (LNGFR), truncated LDL receptor, truncated human growth hormone receptor, GFP, RFP, BFP.

The target gene expression vectors contain a gene (target gene) for expression from the chromosome platform. The target gene can be expressed using various constitutive or regulated promoter systems

- across various mammalian species. For the expression of multiple target genes within the same target gene expression vector, the expression of the multiple targets can be coordinately regulated via viral-based or human internal ribosome entry site (IRES) elements (see, *e.g.*, Jackson *et al.* (1990) *Trends Biochem Sci.* 15: 477-83; Oumard *et al.* (2000) *Mol. Cell. Biol.* 20: 2755-2759). Furthermore, using IRES type elements linked to a downstream fluorescent marker, *e.g.*, green, red or blue fluorescent proteins (GFP, RFP, BFP) allows for the identification of high expressing clones from the integrated target gene expression vector.
- 10 In certain embodiments described herein, the promoterless marker can be transcriptionally downstream of the heterologous nucleic acid, wherein the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.
- 15 In addition, the selectable marker and the heterologous nucleic acid can be transcriptionally linked by the presence of a IRES between them. As set forth herein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or fluorescent.
- 20 Expression from the target gene expression vector integrated onto the chromosome-based platform can be further enhanced using genomic insulator/boundary elements. The incorporation of insulator sequences into the target gene expression vector helps define boundaries in chromatin structure and thus minimizes influence of chromatin position effects/gene silencing on the expression of the target gene (Bell *et al.* (1999) *Current Opinion in Genetics and Development* 9:191-198; Emery *et al.* (2000) *Proc. Natl. Acad. Sci. U.S.A.* 97:9150-9155). Examples of insulator elements that can be included onto target gene expression vector in order to optimize expression include, but are not limited to:
- 25

- 1) chicken  $\beta$ -globin HS4 element (Prioleau *et al.* (1999) *EMBO J* 18: 4035-4048);
- 2) matrix attachment regions (MAR; see, *e.g.*, Ramakrishnan *et al.* (2000) *Mol Cell. Biol.* 20:868-877);
- 5 3) scaffold attachment regions (SAR; see, *e.g.*, Auten *et al.* (1999) *Human Gene Therapy* 10:1389-1399); and
- 4) universal chromatin opening elements (UCOE; WO/0005393 and WO/0224930)

The copy number of the target gene can be controlled by sequentially adding multiple target gene expression vectors containing the target gene onto multiple integration sites on the chromosome platform. Likewise, the copy number of the target gene can be controlled within an individual target gene expression vector by the addition of DNA sequences that promote gene amplification. For example, gene amplification can be induced utilizing the dihydrofolate reductase (DHFR) minigene with subsequent selection with methotrexate (see, *e.g.*, Schimke (1984) *Cell* 37:705-713) or amplification promoting sequences from the rDNA locus (see, *e.g.*, Wegner *et al.* (1989) *Nucl. Acids Res.* 17: 9909-9932).

## 20        6.       Platforms with other recombinase system sites

A "double *lox*" targeting strategy mediated by Cre-recombinase (Bethke *et al.* (1997) *Nucl. Acids Res.* 25:2828-2834) can be used. This strategy employs a pair of heterospecific *lox* sites—*loxA* and *loxB*, which differ by one nucleotide in the 8 bp spacer region. Both sites are engineered into the artificial chromosome and also onto the targeting DNA vector. This allows for a direct site-specific insertion of a commercially relevant gene or genes by a Cre-catalyzed double crossover event. In essence a platform ACes is engineered with a hygromycin-resistance gene flanked by the double *lox* sites generating *lox-ACes*, which is maintained

in the thymidine kinase deficient cell, LMtk(-). The gene of interest, for example, for testing purposes, the green fluorescence protein gene, GFP and a HSV thymidine kinase gene (tk) marker, are engineered between the appropriate *lox* sites of the targeting vector. The vector DNA is  
5 cotransfected with plasmid pBS185 (Life Technologies) encoding the *Cre* recombinase gene into mammalian cells maintaining the dual-*lox* artificial chromosome. Transient expression of the *Cre* recombinase catalyzes the site-specific insertion of the gene and the tk-gene onto the artificial chromosome. The transfected cells are grown in HAT medium that  
10 selects for only those cells that have integrated and expressed the thymidine kinase gene. The HAT<sup>R</sup> colonies are screened by PCR analyses to identify artificial chromosomes with the desired insertion.

To generate the *lox-ACes*, Lambda-Hyg<sup>R</sup>-*lox* DNA is transfected into the LMtk(-) cell line harboring the precursor *ACes*. Hygromycin-  
15 resistant colonies are analyzed by FISH and Southern blotting for the presence of a single copy insert on the *ACes*.

To demonstrate the gene replacement technology, cell lines containing candidate *lox-ACes* are cotransfected with pTK-GFP-*lox* and pBS185 (encoding the *Cre* recombinase gene) DNA. After transfection,  
20 transient expression of plasmid pBS185 will provide sufficient burst of *Cre* recombinase activity to catalyze DNA recombination at the *lox* sites. Thus, a double crossover event between the *ACes* target and the exogenous targeting plasmid carrying the *loxA* and *loxB* permits the simple replacement of the hygromycin-resistance gene on the *lox-ACes*  
25 for the tk-GFP cassette from the targeting plasmid, with no integration of vector DNA. Transfected cells are grown in HAT-media to select for tk-expression. Correct targeting will result in the generation of HAT<sup>R</sup>, hygromycin sensitive, and green fluorescent cells. The desired integration event is verified by Southern and PCR analyses. Specific PCR primer sets

are used to amplify DNA sequences flanking the individual *loxA* and *loxB* sites on the *lox-ACes* before and after homologous recombination.

**D. Exemplary applications of the Platform *ACes***

Platform *ACes* are applicable and tractable for different/optimized  
5 cell lines. Those that include a fluorescent marker, for example, can be  
purified and isolated using fluorescent activated cell sorting (FACS), and  
subsequently delivered to a target cell. Those with selectable markers  
provide for efficient selection and provide a growth advantage. Platform  
*ACes* allow multiple payload delivery of donor target vectors via a  
10 positive-selection site-specific, recombination system, and they allow for  
the inclusion of additional genetic factors that improve protein production  
and protein quality.

The construction and use of the platform *ACes* as provided for  
each application may be similarly applied to other applications. Particular  
15 descriptions are for exemplification.

**1. Cellular Protein Production Platform *ACes* (CPP *ACes*)**

As described herein, *ACes* can be produced from acrocentric  
chromosomes in rodent (mouse, hamster) cell lines via megareplicator  
induced amplification of heterochromatin/rDNA sequences. Such *ACes*  
20 are ideal for cellular protein production as well as other applications  
described herein and known to those of skill in the art. *ACes* platforms  
that contain a plurality of recombination sites are particularly suitable for  
engineering as cellular protein production systems.

In one embodiment, CPP *ACes* involve a two-component system:  
25 the platform chromosome containing multiple engineering sites and the  
donor target vector containing a platform-specific recombination site with  
designed expression cassettes (see Figure 9).

The platform *ACes* can be produced from any artificial  
chromosome, particularly the amplification-based artificial chromosomes.

For exemplification, they are produced from rodent artificial chromosomes produced from acrocentric chromosomes using the technology of U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183, in which nucleic acid is targeted to the pericentric heterochromatic, and, particularly into rDNA to initiate the replication event(s). The ACes can be produced directly in the chosen cellular protein production cell lines, such as, but not limited to, CHO cells, hybridomas, plant cells, plant tissues, plant protoplasts, stem cells and plant calli.

10                   a.       **Platform Construction**

In the exemplary embodiment, the initial *de novo* platform construction requires co-transfecting with excess targeting DNA, such as, rDNA or lambda DNA without an *attP* region, and an engineered selectable marker. The engineered selectable marker should contain promoter, generally a constitutive promoter, such as human, viral, i.e., adenovirus or SV40 promoter, including the human ferritin heavy chain promoter (SEQ ID NO:128), SV40 and EF1 $\alpha$  promoters, to control expression of a marker gene that provides a selective growth advantage to the cell. An example of such a marker gene is the *E. coli hisD* gene (encoding histidinol dehydrogenase) which is homologous and analogous to the *S. typhimurium hisD* a dominant marker selection system for mammalian cells previously described (see, Hartman *et al.* (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85:8047-8051). Since histidine is an essential amino acid in mammals and a nutritional requirement in cell culture, the *E. coli hisD* gene can be used to select for histidine prototrophy in defined media. Furthermore more stringent selection can be placed on the cells by including histinol in the medium. Histidinol is itself permeable and toxic to cells. The *hisD* provides a means of detoxification.



Placed between the promoter and the marker gene is the bacteriophage lambda *attP* site to use the bacteriophage lambda integrase dependent site-specific recombination system (described herein). The insertion of an *attP* site downstream of a promoter element provide  
5 forward selection of site-specific recombination events onto the platform *ACes*.

**b. Donor Target Vector Construction**

A second component of the CPP platform *ACes* system involves the construction of donor target vectors containing a gene product(s) of  
10 interest for the CPP platform *ACes*. Individual donor target vectors can be designed for each gene product to be expressed thus enabling maximum usage of a *de novo* constructed platform *ACes*, so that one or a few CPP platform *ACes* will be required for many gene targets.

A key feature of the donor vector target is the *promoterless* marker  
15 gene containing an upstream *attB* site (marker 2 on figure 9). Normally the marker would not be expressed unless it is placed downstream of a promoter sequence. As discussed above, using the lambda integrase technology ( $\lambda$ INT<sub>E174R</sub> on Figure 8 and Figure 9), site-specific recombination between the *attB* site on the vector and the promoter-*attP*  
20 site on the CPP platform *ACes* result in the expression of the donor target vector marker providing positive selection for the site-specific event. Site-specific recombination events on the CPP *ACes* versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR.  
25 In addition, since the lambda integrase reaction is unidirectional, i.e. excision reaction is not possible, a number of unique targets can be loaded onto the CPP platform *ACes* limited only by the number of markers available.

Additional features of the donor target vector include gene target expression cassettes flanked by either chromatin insulator regions, matrix attachment regions (MAR) or scaffold attachment regions (SAR). The use of these regions will provide a more "open" chromatin environment for gene expression and help alleviate silencing. An example of such a cassette for expressing a monoclonal antibody is described. For this purpose, a strong constitutive promoter, e.g. chicken  $\beta$ -actin or RNA PolI, is used to drive the expression of the heavy and light chain open reading frames. The heavy and light chain sequences flank a nonattenuated human IRES (IRES<sub>H</sub>; from the 5'UTR of NRF1 gene; see Oumard et al., 2000, *Mol. and Cell Biol.*, 20(8):2755-2759) element thereby coordinating transcription of both heavy and light chain sequence. Distal to the light chain open reading frame resides an additional viral encoded IRES (IRES<sub>V</sub> modified ECMV internal ribosomal entry site (IRES)) element attenuating the expression of the fluorescent marker gene hrGFP from *Renilla* (Stratagene). By linking the hrGFP with an attenuated IRES, the heavy and light chains along with the hrGFP are monocistronic. Thus, the identification of hrGFP fluorescing cells will provide a means to detect protein producing cells. In addition, high producing cell lines can be identified and isolated by FACS thereby decreasing the time frame in finding high expressers. Functional monoclonal antibody will be confirmed by ELISA.

**c. Additional components in cellular protein production platform ACes (CPP ACes)**

In addition to the aforementioned CPP ACes system, other genetic factors can be included to enhance the yield and quality of the expressed protein. Again to provide maximum flexibility, these additional factors can be inserted onto the CPP platform ACes by  $\lambda$ INTE174R dependent site-specific recombination. Other factors that could be used with a CPP

Platform ACes include for example, adenovirus E1a transactivation system which upregulates both cellular and viral promoters (see, e.g., Svensson and Akusjarvi (1984) EMBO 3:789-794; and US patents 5,866,359; 4,775,630 and 4,920,211).

5                   d.       **Targets for CHO-ACes engineering to enhance cell growth, such as CHO cell growth and protein production/ quality**

                  If adding these additional factors onto the CPP ACes is not prudent or desired, the host cell, CHO cells, can be engineered to express these factors (see, below, targets for CHO-ACes engineering to enhance CHO cell growth and protein production/quality). Additional factors to consider including are addition of insulin or IGF-1 to sustain viability; human sialyltransferases or related factors to produce more human-like glycoproteins; expression of factors to decrease ammonium accumulation during cell growth; expression of factors to inhibit apoptosis; expression of factors to improve protein secretion and protein folding; and expression of factors to permit serum-free transfection and selection.

                                  1)       **Addition of insulin or IGF-1 to sustain viability**

20               Stimulatory factors and/or their receptors are expressed to set up an autocrine loop, to improve cell growth, such as CHO cell growth. Two exemplary candidates are insulin and IGF-1 (see, Biotechnol Prog 2000 Sep;16(5):693-7). Insulin is the most commonly used growth factor for sustaining cell growth and viability in serum-free Chinese hamster ovary (CHO) cell cultures. Insulin and IGF-1 analog (LongR(3) serve as growth and viability factors for CHO cells.

                  CHO cells were modified to produce higher levels of essential nutrients and factors. A serum-free (SF) medium for dihydrofolate reductase-deficient Chinese hamster ovary cells (DG44 cells) was prepared. Chinese hamster ovary cells (DG44 cells), which are normally

maintained in 10% serum medium, were gradually weaned to 0.5% serum medium to increase the probability of successful growth in SF medium (see, Kim *et al.* (199) *In Vitro Cell Dev Biol Anim* 35(4):178-82). A SF medium (SF-DG44) was formulated by supplementing the basal medium with these components; basal medium was prepared by supplementing Dulbecco's modified Eagle's medium and Ham's nutrient mixture F12 with hypoxanthine (10 mg/l) and thymidine (10 mg/l). Development of a SF medium for DG44 cells was facilitated using a Plackett-Burman design technique and weaning of cells.

10

**2) Human sialyltransferases or related factors to produce more human-like glycoproteins**

CHO cells have been modified by increasing their ability to process protein via addition of complex carbohydrates. This has been achieved by overexpression of relevant processing enzymes, or in some cases, reducing expression of relevant enzymes (see, Bragonzi *et al.* (2000) *Biochim Biophys Acta* 1474(3):273-282; see, also Weikert *et al.* (1999) *Nature biotech.* 17:1116-11121; Ferrari J *et al.* (1998) *Biotechnol Bioeng* 60(5):589-95). A CHO cell line expressing alpha2,6-sialyltransferase was developed for the production of human-like sialylated recombinant glycoproteins. The sialylation defect of CHO cells can be corrected by transfecting the alpha2,6-sialyltransferase (alpha2,6-ST) cDNA into the cells. Glycoproteins produced by such CHO cells display alpha2,6-and alpha2,3-linked terminal sialic acid residues, similar to human glycoproteins.

As another example for improving the production of human-like sialylated recombinant glycoproteins, a CHO cell line has been developed that constitutively expresses sialidase antisense RNA (see, Ferrari J *et al.* (1998) *Biotechnol Bioeng* 60(5):589-95). Several antisense expression

30

vectors were prepared using different regions of the sialidase gene. Co-transfection of the antisense constructs with a vector conferring puromycin resistance gave rise to over 40 puromycin resistant clones that were screened for sialidase activity. A 5' 474 bp coding segment of the sialidase cDNA, in the inverted orientation in an SV 40-based expression vector, gave maximal reduction of the sialidase activity to about 40% wild-type values.

Oligosaccharide biosynthesis pathways in mammalian cells have been engineered for generation of recombinant glycoproteins (see, *e.g.*, Sburlati (1998) *Biotechnol Prog* 14(2):189-92), which describes a Chinese hamster ovary (CHO) cell line capable of producing bisected oligosaccharides on glycoproteins. This cell line was created by overexpression of a recombinant N-acetylglucosaminyltransferase III (GnT-III) (see, also, Prati *et al.* (1998) *Biotechnol Bioeng* 59(4):445-50, which describes antisense strategies for glycosylation engineering of CHO cells).

### **3) Expression of factors to decrease ammonium accumulation during cell growth**

Excess ammonium, which is a by-product of CHO cell metabolism can have detrimental effects on cell growth and protein quality (see, Yang *et al.* (2000) *Biotechnol Bioeng* 68(4):370-80). To solve this problem ammonium levels were modified by overexpressing carbamoyl phosphate synthetase I and ornithine transcarbamoylase or glutamine synthetase in CHO cells. Such modification resulted in reduced ammonium levels observed and an increase in the growth rate (see Kim *et al.* (2000) *J Biotechnol* 81(2-3):129-40; and Enosawa *et al.* (1997) *Cell Transplant* 6(5):537-40).

### **4) Expression of factors to improve protein secretion and protein folding**

Overexpression of relevant enzymes can be engineered into the ACes to improve protein secretion and folding.

**5) Expression of factors to permit serum-free transfection and selection**

5 It is advantageous to have the ability to convert CHO cells in suspension growing in serum free medium to adherence without having to resort to serum addition. Laminin or fibronectin addition is sufficient to make cells adherent (see, *e.g.*, Zaworski *et al.* (1993) *Biotechniques* 15(5):863-6) so that expressing either of these genes in CHO cells under  
10 an inducible promoter should allow for reversible shift to adherence without requiring serum addition.

**2. Platform ACes and Gene Therapy**

The platform ACes provided herein are contemplated for use in mammalian gene therapy, particularly human gene therapy. Human ACes  
15 can be derived from human acrocentric chromosomes from human host cells, in which the amplified sequences are heterochromatic and/or human rDNA. Different platform ACes applicable for different tissue cell types are provided. The ACes for gene therapy can contain a single copy of a therapeutic gene inserted into a defined location on platform ACes.

20 Therapeutic genes include genomic clones, cDNA, hybrid genes and other combinations of sequences. Preferred selectable markers are those from the mammalian host, such as human derived factors so that they are non-immunogenic, non-toxic and allow for efficient selection, such as by FACS and/or drug resistance.

25 Platform ACes, useful for gene therapy and other applications, as noted herein, can be generated by megareplicator dependent amplification, such as by the methods in U.S. Patent Nos. 6,077,697 and 6,025,155 and published International PCT application No. WO 97/40183. In one embodiment, human ACes are produced using

human rDNA constructs that target rDNA arrays on human acrocentric chromosomes and induce the megareplicator in human cells, particularly in primary cell lines (with sufficient number of doublings to form the ACes) or stem cells (such as hematopoietic stem cells, mesenchymal stem cells, adult stem cells or embryonic stem cells) to avoid the introduction of potentially harmful rearranged DNA sequences present in many transformed cell lines. Megareplicator induced ACes formation can result in multiple copies of targeting DNA/selectable markers in each amplification block on both chromosomal arms of the platform ACes.

10 In view of the considerations regarding immunogenicity and toxicity, the production of human platform ACes for gene therapy applications employs a two component system analogous to the platform ACes designed for cellular protein production (CPP platform ACes). The system includes a platform chromosome of entirely human DNA origin  
15 containing multiple engineering sites and a gene target vector carrying the therapeutic gene of interest.

**a. Platform Construction**

The initial *de novo* construction of the platform chromosome employs the co-transfection of excess targeting DNA and a selectable  
20 marker. In one embodiment, the DNA is targeted to the rDNA arrays on the human acrocentric chromosomes (chromosomes 13, 14, 15, 21 and 22). For example, two large human rDNA containing PAC clones 18714 and 18720 and the human PAC clone 558F8 are used for targeting (Genome Research (ML) now Incyte, BACPAC Resources, 747 52nd  
25 Street, Oakland CA). The mouse rDNA clone pFK161 (SEQ ID NO: 118), which was used to make the human SATAC from the 94-3 hamster/human hybrid cell line (see, *e.g.*, published International PCT application No. WO 97/40183 and Csonka, *et al*, *Journal of Cell Science*

113:3207-32161 and Example 1 for a description of pFK161) can also be used.

For animal applications, selectable markers should be non-immunogenic in the animal, such as a human, and include, but are not limited to: human nerve growth factor receptor (detected with a MAb, such as described in US patent 6,365,373); truncated human growth factor receptor (detected with MAb), mutant human dihydrofolate reductase (DHFR; fluorescent MTX substrate available); secreted alkaline phosphatase (SEAP; fluorescent substrate available); human thymidylate synthase (TS; confers resistance to anti-cancer agent fluorodeoxyuridine); human glutathione S-transferase alpha (GSTA1; conjugates glutathione to the stem cell selective alkylator busulfan; chemoprotective selectable marker in CD34 + cells); CD24 cell surface antigen in hematopoietic stem cells; human CAD gene to confer resistance to N-phosphonacetyl-L-aspartate (PALA); human multi-drug resistance-1 (MDR-1; P-glycoprotein surface protein selectable by increased drug resistance or enriched by FACS); human CD25 (IL-2 $\alpha$ ; detectable by Mab-FITC); Methylguanine-DNA methyltransferase (MGMT; selectable by carmustine); and Cytidine deaminase (CD; selectable by Ara-C).

Since megareplicator induced amplification generates multiple copies of the selectable marker, a second consideration for the selection of the human marker is the resulting dose of the expressed marker after ACes formation. High level of expression of certain markers may be detrimental to the cell and/or result in autoimmunity. One method to decrease the dose of the marker protein is by shortening its half-life, such as via the fusion of the well-conserved human ubiquitin tag (a 76 amino acid sequence) thus leading to increased turnover of the selectable marker. This has been used successfully for a number of reporter



systems including DHFR (see, *e.g.*, Stack *et al.* (2000) *Nature Biotechnology* 18:1298-1302 and references cited therein).

Using the ubiquitin tagged protein, a human selectable marker system analogous to the CPP ACes described herein is constructed.

- 5 Briefly, a tagged selectable marker, such as for example one of those described herein, is cloned downstream of an *attP* site and expressed from a human promoter. Exemplary promoters contemplated for use herein include, but are not limited to, the human ferritin heavy chain promoter (SEQ ID NO:128); RNA Poll; EF1 $\alpha$ ; TR; glyceraldehyde-3-  
10 phosphate dehydrogenase core promoter (GAP); a GAP core promoter including a proximal insulin inducible element and the intervening GAP sequence; phosphofructokinase promoter; and phosphoglycerate kinase promoter. Also contemplated herein is an aldolase A promoter H1 & H2 (representing closely spaced transcriptional start sites) along with the  
15 proximal H enhancer. There are 4 promoters (*e.g.*, transcriptional start sites) for this gene, each having different regulatory and tissue activity. The H (most proximal 2) promoters are ubiquitously expressed off the H enhancer. This resulting marker can then be co-transfected along with excess human rDNA targeting sequence into the host cells. An important  
20 criteria for the selection of the

- recipient cells is sufficient number of cell doublings for the formation and detection of ACes. Accordingly, the co-transfections should be attempted in human primary cells that can be cultured for long periods of  
25 time, such as for example, stem cells (*e.g.*, hematopoietic, mesenchymal, adult or embryonic stem cells), or the like. Additional cell types, include, but are not limited to: single gene transfected cells exhibiting increased life-span; over-expressing c-myc cells, *e.g.* MSU1.1 (Morgan *et al.*, 1991, *Exp. Cell Res.*, Nov;197(1):125-136); over-expressing telomerase lines,

such as TERT cells; SV40 large T-antigen transfected lines; tumor cell lines, such as HT1080; and hybrid human cell lines, such as the 94-3 hamster/human hybrid cell line.

**b. Gene Target Vector**

5           The second component of the GT platform *ACes* (GT *ACes*) system involves the use of engineered target vectors carrying the therapeutic gene of interest. These are introduced onto the GT platform *ACes* via site-specific recombination. As with the CPP *ACes*, the use of engineered target vectors maximizes the use of the *de novo* generated GT platform  
10 *ACes* for most gene targets. Furthermore, using lambda integrase technology, GT platform *ACes* containing multiple *attP* sites permits the opportunity to incorporate multiple therapeutic targets onto a single platform. This could be of value in cases where a defined therapy requires multiple gene targets, a single therapeutic target requires an  
15 additional gene regulatory factor or a GT *ACes* requires a "kill" switch.

          Similar to the CPP *ACes*, a feature of the gene target vector is the *promoterless* marker gene containing an upstream *attB* site (marker 2 on Figure 9). Normally, the marker (in this case, a cell surface antigen that can be sorted by FACS would be ideal) would not be expressed unless it  
20 is placed downstream of a promoter sequence. Using the lambda integrase technology ( $\lambda$ INT<sub>E174R</sub> on figure 9), site-specific recombination between the *attB* site on the vector and the promoter- *attP* site on the GT platform *ACes* results in the expression of marker#2 on the gene target vector, i.e. positive selection for the site-specific event. Site-specific  
25 recombination events on the GT *ACes* versus random integrations next to a promoter in the genome (false positive) can be quickly screened by designing primers to detect the correct event by PCR.

          For expression of the therapeutic gene, human specific promoters, such as a ferritin heavy chain promoter (SEQ ID NO:128); EF1 $\alpha$  or RNA

Poll, are used. These promoters are for high level expression of a cDNA encoded therapeutic protein. In addition to expressing cDNA (or even hybrid cDNA/artificial intron constructs), the GT platform ACes are used for engineering and expressing large genomic fragments carrying  
5 therapeutic genes of interest expressed from native promoter sequences. This is of importance in situations where the therapy requires precise cell specific expression or in instances where expression is best achieved from genomic clones versus cDNA.

10                   **3.     Selectable markers for use, for example, in Gene Therapy (GT)**

          The following are selectable markers that can be incorporated into human ACes and used for selection.

15                               **Dual Resistance to 4-Hydroperoxycyclophosphamide and Methotrexate by Retroviral Transfer of the Human Aldehyde Dehydrogenase Class 1 Gene and a Mutated Dihydrofolate Reductase Gene**

          The genetic transfer of drug resistance to hematopoietic cells is one approach to overcoming myelosuppression caused by high-dose chemotherapy. Because cyclophosphamide (CTX) and methotrexate  
20 (MTX) are commonly used non-cross-resistant drugs, generation of dual drug resistance in hematopoietic cells that allows dose intensification may increase anti-tumor effects and circumvent the emergence of drug-resistant tumors, a retroviral vector containing a human cytosolic ALDH-1-encoding DNA clone and a human doubly mutated DHFR-encoding  
25 clone (Phe22/Ser31; termed F/S in the description of constructs) to generate increased resistance to CTX and MTX were constructed (Takebe *et al.* (2001) *Mol Ther* 3(1):88-96). This construct may be useful for protecting patients from high-dose CTX- and MTX-induced myelosuppression. ACes can be similarly constructed.

**Multiple mechanisms of N-phosphonacetyl-L-aspartate resistance in human cell lines: carbamyl-P synthetase/aspartate transcarbamylase/dihydro-orotase gene amplification is frequent only when chromosome 2 is rearranged**

5 Rodent cells resistant to N-phosphonacetyl-L-aspartate (PALA) invariably contain amplified carbamyl-P synthetase/aspartate transcarbamylase/dihydro-orotase (CAD) genes, usually in widely spaced tandem arrays present as extensions of the same chromosome arm that  
10 carries a single copy of CAD in normal cells (Smith *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:1816-21). In contrast, amplification of CAD is very infrequent in several human tumor cell lines. Cell lines with minimal chromosomal rearrangement and with unrearranged copies of chromosome 2 rarely develop intrachromosomal amplifications of CAD.  
15 These cells frequently become resistant to PALA through a mechanism that increases the aspartate transcarbamylase activity with no increase in CAD copy number, or they obtain one extra copy of CAD by forming an isochromosome 2p or by retaining an extra copy of chromosome 2. In cells with multiple chromosomal aberrations and rearranged copies of  
20 chromosome 2, amplification of CAD as tandem arrays from rearranged chromosomes is the most frequent mechanism of PALA resistance. All of these different mechanisms of PALA resistance are blocked in normal human fibroblasts. Thus, ACes with multiple copies of the CAD gene would provide PALA resistance.

25 **Retroviral coexpression of thymidylate synthase and dihydrofolate reductase confers fluoropyrimidine and antifolate resistance**

Retroviral gene transfer of dominant selectable markers into hematopoietic cells can be used to select genetically modified cells in vivo  
30 or to attenuate the toxic effects of chemotherapeutic agents. Fantz *et al.* ((1998) *Biochem Biophys Res Comm* 243(1):6-12) have shown that

retroviral gene transfer of thymidylate synthase (TS) confers resistance to TS directed anticancer agents and that co-expression of TS and dihydrofolate reductase (DHFR) confers resistance to TS and DHFR cytotoxic agents. Retroviral vectors encoding *Escherichia coli* TS, human TS, and the Tyr-to-His at residue 33 variant of human TS (Y33HhTS) were constructed and fibroblasts transfected with these vectors conferred comparable resistance to the TS-directed agent fluorodeoxyuridine (FdUrd, approximately 4-fold). Retroviral vectors that encode dual expression of Y33HhTS and the human L22Y DHFR (L22YhDHFR) variants conferred resistance to FdUrd (3- to 5-fold) and trimetrexate (30- to 140-fold). A L22YhDHFR-Y33HhTS chimeric retroviral vector was also constructed and transduced cells were resistant to FdUrd (3-fold), AG337 (3-fold), trimetrexate (100-fold) and methotrexate (5-fold). These results show that recombinant retroviruses can be used to transfer the cDNA that encodes TS and DHFR and dual expression in transduced cells is sufficiently high to confer resistance to TS and DHFR directed anticancer agents. ACes can be similarly constructed.

**Human CD34+ cells do not express glutathione S-transferases alpha**

The expression of glutathione S-transferases alpha (GST alpha) in human hematopoietic CD34+ cells and bone marrow was studied using RT-PCR and immunoblotting (Czerwinski M, Kiem *et al.* (1997) *Gene Ther* 4(3):268-70). The GSTA1 protein conjugates glutathione to the stem cell selective alkylator busulfan. This reaction is the major pathway of elimination of the compound from the human body. Human hematopoietic CD34+ cells and bone marrow do not express GSTA1 message, which was present at a high level in liver, an organ relatively resistant to busulfan toxicity in comparison to bone marrow. Similarly, baboon CD34+ cells and dog bone marrow do not express GSTA1. Thus, human

GSTA1 is a chemoprotective selectable marker in human stem cell gene therapy and could be employed in ACes construction.

**Selection of retrovirally transduced hematopoietic cells using CD24 as a marker of gene transfer**

- 5 Pawliuk *et al.* ((1994) *Blood* 84(9):2868-2877) have investigated the use of a cell surface antigen as a dominant selectable marker to facilitate the detection and selection of retrovirally infected target cells. The small coding region of the human cell surface antigen CD24 (approximately 240 bp) was introduced into a myeloproliferative sarcoma virus (MPSV)-based retroviral vector, which was then used to infect day 4
- 10 5-fluorouracil (5-FU)-treated murine bone marrow cells. Within 48 hours of termination of the infection procedure CD24-expressing cells were selected by fluorescent-activated cell sorting (FACS) with an antibody directed against the CD24 antigen. Functional analysis of these cells
- 15 showed that they included not only *in vitro* clonogenic progenitors and day 12 colony-forming unit-spleen but also cells capable of competitive long-term hematopoietic repopulation. Double-antibody labeling studies performed on recipients of retrovirally transduced marrow cells showed that some granulocytes, macrophages, erythrocytes, and, to a lesser
- 20 extent, B and T lymphocytes still expressed the transduced CD24 gene at high levels 4 months later. No gross abnormalities in hematopoiesis were detected in mice repopulated with CD24-expressing cells. These results show that the use of the CD24 cell surface antigen as a retrovirally encoded marker permits rapid, efficient, and nontoxic selection *in vitro* of
- 25 infected primary cells, facilitates tracking and phenotyping of their progeny, and provides a tool to identify elements that regulate the expression of transduced genes in the most primitive hematopoietic cells. ACes could be similarly constructed.

**DeltahGHR, a biosafe cell surface-labeling molecule for analysis and selection of genetically transduced human cells**

A selectable marker for retroviral transduction and selection of human and murine cells is known (see, Garcia-Ortiz *et al.* (2000) *Hum Gene Ther* 11(2):333-46). The molecule expressed on the cell surface of the transduced population is a truncated version of human growth hormone receptor (deltahGHR), capable of ligand (hGH) binding, but devoid of the domains involved in signal triggering. The engineered molecule is stably expressed in the target cells as an inert protein unable to trigger proliferation or to rescue the cells from apoptosis after ligand binding. This new marker, has a wide application spectrum, since hGHR in the human adult is highly expressed only in liver cells, and lower levels have been reported in certain lymphocyte cell populations. The deltagHHR label has high biosafety potential, as it belongs to a well-characterized hormonal system that is nonessential in adults, and there is extensive clinical experience with hGH administration in humans. The differential binding properties of several monoclonal antibodies (MAbs) are used in a cell rescue method in which the antibody used to select deltagHHR-transduced cells is eluted by competition with hGH or, alternatively biotinylated hGH is used to capture tagged cells. In the latter system, the final purified population is recovered free of attached antibodies in hGH (a substance approved for human use)-containing medium. Such a system could be used to identify ACes containing cells.

**4. Transgenic models for evaluation of genes and discovery of new traits in plants**

Of interest is the use of plants and plant cells containing artificial chromosomes for the evaluation of new genetic combinations and discovery of new traits. Artificial chromosomes, by virtue of the fact that they can contain significant amounts of DNA can also therefore encode

numerous genes and accordingly a multiplicity of traits. It is contemplated here that artificial chromosomes, when formed from one plant species, can be evaluated in a second plant species. The resultant phenotypic changes observed, for example, can indicate the nature of the genes contained within the DNA contained within the artificial chromosome, and hence permit the identification of novel genetic activities. Artificial chromosomes containing euchromatic DNA or partially containing euchromatic DNA can serve as a valuable source of new traits when transferred to an alien plant cell environment. For example, it is contemplated that artificial chromosomes derived from dicot plant species can be introduced into monocot plant species by transferring a dicot artificial chromosome. The dicot artificial chromosome possessing a region of euchromatic DNA containing expressed genes.

The artificial chromosomes can be designed to allow the artificial chromosome to recombine with the naturally occurring plant DNA in such a fashion that a large region of naturally occurring plant DNA becomes incorporated into the artificial chromosome. This allows the artificial chromosome to contain new genetic activities and hence carry novel traits. For example, an artificial chromosome can be introduced into a wild relative of a crop plant under conditions whereby a portion of the DNA present in the chromosomes of the wild relative is transferred to the artificial chromosome. After isolation of the artificial chromosome, this naturally occurring region of DNA from the wild relative, now located on the artificial chromosome can be introduced into the domesticated crop species and the genes encoded within the transferred DNA expressed and evaluated for utility. New traits and gene systems can be discovered in this fashion. The artificial chromosome can be modified to contain sequences that promote homologous recombination within plant cells, or



be modified to contain a genetic system that functions as a site-specific recombination system.

Artificial chromosomes modified to recombine with plant DNA offer many advantages for the discovery and evaluation of traits in different  
5 plant species. When the artificial chromosome containing DNA from one plant species is introduced into a new plant species, new traits and genes can be introduced. This use of an artificial chromosome allows for the ability to overcome the sexual barrier that prevents transfer of genes from one plant species to another species. Using artificial chromosomes in this  
10 fashion allows for many potentially valuable traits to be identified including traits that are typically found in wild species. Other valuable applications for artificial chromosomes include the ability to transfer large regions of DNA from one plant species to another, such as DNA encoding potentially valuable traits such as altered oil, carbohydrate or protein  
15 composition, multiple genes encoding enzymes capable of producing valuable plant secondary metabolites, genetic systems encoding valuable agronomic traits such as disease and insect resistance, genes encoding functions that allow association with soil bacterium such as growth promoting bacteria or nitrogen fixing bacteria, or genes encoding traits  
20 that confer freezing, drought or other stress tolerances. In this fashion, artificial chromosomes can be used to discover regions of plant DNA that encode valuable traits.

The artificial chromosome can also be designed to allow the transfer and subsequent incorporation of these valuable traits now located  
25 on the artificial chromosome into the natural chromosomes of a plant species. In this fashion the artificial chromosomes can be used to transfer large regions of DNA encoding traits normally found in one plant species into another plant species. In this fashion, it is possible to derive a plant cell that no longer needs to carry an artificial chromosome to

posses the novel trait. Thus, the artificial chromosome would serve as the transfer mechanism to permit the formation of plants with greater degree of genetic diversity.

5 The design of an artificial chromosome to accomplish the afore-  
mentioned purposes can include within the artificial chromosome the  
presence of specific DNA sequences capable of acting as sites for  
homologous recombination to take place. For example, the DNA  
sequence of *Arabidopsis* is now known. To construct an artificial  
chromosome capable of recombining with a specific region of *Arabidopsis*  
10 DNA, a sequence of *Arabidopsis* DNA, normally located near a  
chromosomal location encoding genes of potential interest can be  
introduced into an artificial chromosome by methods provided herein. It  
may be desirable to include a second region of DNA within the artificial  
chromosome that provides a second flanking sequence to the region  
15 encoding genes of potential interest, to promote a double recombination  
event which would ensure transfer of the entire chromosomal region,  
encoding genes of potential interest, to the artificial chromosome. The  
modified artificial chromosome, containing the DNA sequences capable of  
homologous recombination region, can then be introduced into  
20 *Arabidopsis* cells and the homologous recombination event selected.

It is convenient to include a marker gene to allow for the selection  
of a homologous recombination event. The marker gene is preferably  
inactive unless activated by an appropriate homologous recombination  
event. For example, US 5,272,071, describes a method where an  
25 inactive plant gene is activated by a recombination event such that  
desired homologous recombination events can be easily scored. Similarly,  
US 5,501,967 describes a method for the selection of homologous  
recombination events by activation of a silent selection gene first  
introduced into the plant DNA, the gene being activated by an appropriate

homologous recombination event. Both of these methods can be applied to enable a selective process to be included to select for recombination between an artificial chromosome and a plant chromosome. Once the homologous recombination event is detected, the artificial chromosome, 5 once selected, is isolated and introduced into a recipient cell, for example, tobacco, corn, wheat or rice, and the expression of the newly introduced DNA sequences evaluated.

Phenotypic changes in the recipient plant cells containing the artificial chromosome, or in regenerated plants containing the artificial 10 chromosome, allows for the evaluation of the nature of the traits encoded by the *Arabidopsis* DNA, under conditions naturally found in plant cells, including the naturally occurring arrangement of DNA sequences responsible for the developmental control of the traits in the normal chromosomal environment.

15 Traits such as durable fungal or bacterial disease resistance, new oil and carbohydrate compositions, valuable secondary metabolites such as phytosterols, flavonoids, efficient nitrogen fixation or mineral utilization, resistance to extremes of drought, heat or cold are all found within different populations of plant species and are often governed by 20 multiple genes. The use of single gene transformation technologies does not permit the evaluation of the multiplicity of genes controlling many valuable traits. Thus, incorporation of these genes into artificial chromosomes allows the rapid evaluation of the utility of these genetic combinations in heterologous plant species.

25 The large scale order and structure of the artificial chromosome provides a number of unique advantages in screening for new utilities or novel phenotypes within heterologous plant species. The size of new DNA that can be carried by an artificial chromosome can be millions of base pairs of DNA, representing potentially numerous genes that may

have novel utility in a heterologous plant cell. The artificial chromosome is a "natural" environment for gene expression, the problems of variable gene expression and silencing seen for genes transferred by random insertion into a genome should not be observed. Similarly, there is no  
5 need to engineer the genes for expression, and the genes inserted would not need to be recombinant genes. Thus, one expects the expression from the transferred genes to be temporal and spatial, as observed in the species from where the genes were initially isolated. A valuable feature for these utilities is the ability to isolate the artificial chromosomes and to  
10 further isolate, manipulate and introduce into other cells artificial chromosomes carrying unique genetic compositions.

Thus, the use of artificial chromosomes and homologous recombination in plant cells can be used to isolate and identify many valuable crop traits.

15 In addition to the use of artificial chromosomes for the isolation and testing of large regions of naturally occurring DNA, methods for the use of artificial chromosomes and cloned DNA are also contemplated. Similar to that described above, artificial chromosomes can be used to carry large regions of cloned DNA, including that derived from other plant species.

20 The ability to incorporate novel DNA elements into an artificial chromosome as it is being formed allows for the development of artificial chromosomes specifically engineered as a platform for testing of new genetic combinations, or "genomic" discoveries for model species such as *Arabidopsis*. It is known that specific "recombinase" systems can be  
25 used in plant cells to excise or re-arrange genes. These same systems can be used to derive new gene combinations contained on an artificial chromosome.

The artificial chromosomes can be engineered as platforms to accept large regions of cloned DNA, such as that contained in Bacterial

Artificial Chromosomes (BACs) or Yeast Artificial Chromosomes (YACs). It is further contemplated, that as a result of the typical structure of artificial chromosomes containing tandemly repeated DNA blocks, that sequences other than cloned DNA sequence can be introduced by  
5 recombination processes. In particular recombination within a predefined region of the tandemly repeated DNA within the artificial chromosome provides a mechanism to "stack" numerous regions of cloned DNA, including large regions of DNA contained within BACs or YACs clones. Thus, multiple combinations of genes can be introduced onto artificial  
10 chromosomes and these combinations tested for functionality. In particular, it is contemplated that multiple YACs or BACs can be stacked onto an artificial chromosomes, the BACs or YACs containing multiple genes of complex pathways or multiple genetic pathways. The BACs or YACs are typically selected based on genetic information available within  
15 the public domain, for example from the Arabidopsis Information Management System (<http://aims.cps.msu.edu/aims/index.html>) or the information related to the plant DNA sequences available from the Institute for Genomic Research (<http://www.tigr.org>) and other sites known to those skilled in the art. Alternatively, clones can be chosen at  
20 random and evaluated for functionality. It is contemplated that combinations providing a desired phenotype can be identified by isolation of the artificial chromosome containing the combination and analyzing the nature of the inserted cloned DNA.

In this regard, it is contemplated that the use of site-specific  
25 recombination sequences can have considerable utility in developing artificial chromosomes containing DNA sequences recognized by recombinase enzymes and capable of accepting DNA sequences containing same. The use of site-specific recombination as a means to target an introduced DNA to a specific locus has been demonstrated in

the art and such methods can be employed. The recombinase systems can also be used to transfer the cloned DNA regions contained within the artificial chromosome to the naturally occurring plant or mammalian chromosomes.

5           As noted herein, many site-specific recombinases are known and can be identified (Kilby *et al.* (1993) *Trends in Genetics* 9:413-418). The three recombinase systems that have been extensively employed include: an activity identified as R encoded by the pSR1 plasmid of *Zygosaccharomyces rouxii*, FLP encoded for the 2µm circular plasmid from  
10 *Saccharomyces cerevisiae* and *Cre-lox* from the phage P1.

          The integration function of site-specific recombinases is contemplated as a means to assist in the derivation of genetic combinations on artificial chromosomes. In order to accomplish this, it is contemplated that a first step of introducing site-specific recombinase  
15 sites into the genome of a plant cell in an essentially random manner is conducted, such that the plant cell has one or more site-specific recombinase recognition sequences on one or more of the plant chromosomes. An artificial chromosome is then introduced into the plant cell, the artificial chromosome engineered to contain a recombinase  
20 recognition site (e.g., integration site) capable of being recognized by a site-specific recombinase. Optionally, a gene encoding a recombinase enzyme is also included, preferably under the control of an inducible promoter. Expression of the site-specific recombinase enzyme in the plant cell, either by induction of a inducible recombinase gene, or  
25 transient expression of a recombinase sequence, causes a site-specific recombination event to take place, leading to the insertion of a region of the plant chromosomal DNA (containing the recombinase recognition site) into the recombinase recognition site of the artificial chromosome, and forming an artificial chromosome containing plant chromosomal DNA.

The artificial chromosome can be isolated and introduced into a heterologous host, preferably a plant host, and expression of the newly introduced plant chromosomal DNA can be monitored and evaluated for desirable phenotypic changes. Accordingly, carrying out this

5 recombination with a population of plant cells wherein the chromosomally located recombinase recognition site is randomly scattered throughout the chromosomes of the plant, can lead to the formation of a population of artificial chromosomes, each with a different region of plant chromosomal DNA, and each potentially representing a novel genetic combination.

10 This method requires the precise site-specific insertion of chromosomal DNA into the artificial chromosome. This precision has been demonstrated in the art. For example, Fukushige and Sauer ((1992) Proc. Natl. Acad. Sci. USA, 89:7905-7909) demonstrated that the *Cre-lox* homologous recombination system could be successfully employed to  
15 introduce DNA into a predefined locus in a chromosome of mammalian cells. In this demonstration a promoter-less antibiotic resistance gene modified to include a *lox* sequence at the 5' end of the coding region was introduced into CHO cells. Cells were re-transformed by electroporation with a plasmid that contained a promoter with a *lox* sequence and a  
20 transiently expressed *Cre* recombinase gene. Under the conditions employed, the expression of the *Cre* enzyme catalyzed the homologous recombination between the *lox* site in the chromosomally located promoter-less antibiotic resistance gene, and the *lox* site in the introduced promoter sequence, leading to the formation of a functional antibiotic  
25 resistance gene. The authors demonstrated efficient and correct targeting of the introduced sequence, 54 of 56 lines analyzed corresponded to the predicted single copy insertion of the DNA due to *Cre* catalyzed site-specific homologous recombination between the *lox* sequences.

Accordingly a *lox* sequence may be first added to a genome of a plant species capable of being transformed and regenerated to a whole plant to serve as a recombinase target DNA sequence for recombination with an artificial chromosome. The *lox* sequence may be optimally  
5 modified to further contain a selectable marker which is inactive but can be activated by insertion of the *lox* recombinase recognition sequence into the artificial chromosome.

A promoterless marker gene or selectable marker gene linked to the recombinase recognition sequence, which is first inserted into the  
10 chromosomes of a plant cell can be used to engineer a platform chromosome. A promoter is linked to a recombinase recognition site, in an orientation that allows the promoter to control the expression of the marker or selectable marker gene upon recombination within the artificial chromosome. Upon a site-specific recombination event between a  
15 recombinase recognition site in a plant chromosome and the recombinase recognition site within the introduced artificial chromosome, a cell is derived with a recombined artificial chromosome, the artificial chromosome containing an active marker or selectable marker activity that permits the identification and or selection of the cell.

20 The artificial chromosomes can be transferred to other plant or animal species and the functionality of the new combinations tested. The ability to conduct such an inter-chromosomal transfer of sequences has been demonstrated in the art. For example, the use of the *Cre-lox* recombinase system to cause a chromosome recombination event  
25 between two chromatids of different chromosomes has been shown.

Any number of recombination systems may be employed as described herein, such as, but not limited to, bacterially derived systems such as the att/int system of phage lambda, and the Gin/gix system.



More than one recombination system may be employed, including, for example, one recombinase system for the introduction of DNA into an artificial chromosome, and a second recombinase system for the subsequent transfer of the newly introduced DNA contained within an artificial chromosome into the naturally occurring chromosome of a second plant species. The choice of the specific recombination system used will be dependent on the nature of the modification contemplated.

By having the ability to isolate an artificial chromosome, in particular, artificial chromosomes containing plant chromosomal DNA introduced via site-specific recombination, and re-introduce the chromosome into other mammalian or plant cells, particularly plant cells, these new combinations can be evaluated in different crop species without the need to first isolate and modify the genes, or carry out multiple transformations or gene transfers to achieve the same combination isolation and testing combinations of the genes in plants. The use of a site-specific recombinase also allows the convenient recovery of the plant chromosomal region into other recombinant DNA vectors and systems, such as mammalian or insect systems, for manipulation and study.

Also contemplated herein are *ACes*, cell lines and methods for use in screening a new chromosomal combinations, deletions, truncations with eucaryotic genome that take advantage of the site-specific recombination systems incorporated onto platform *ACes* provided herein. For example, provided herein is a cell line useful for making a library of *ACes*, comprising a multiplicity of heterologous recombination sites randomly integrated throughout the endogenous chromosomes. Also provided herein is a method of making a library of *ACes* comprising random portions of a genome, comprising introducing one or more *ACes* into a cell line comprising a multiplicity of heterologous recombination

sites randomly integrated throughout the endogenous chromosomes, under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites within the cell's chromosomal DNA; and isolating said  
5 multiplicity of *ACes*, thereby producing a library of *ACes* whereby multiple *ACes* have different portions of the genome within. Also provided herein is a library of cells useful for genomic screening, said library comprising a multiplicity of cells, wherein each cell comprises an *ACes* having a mutually exclusive portion of a chromosomal nucleic acid therein. The  
10 library of cells can be from a different species and/or cell type than the chromosomal nucleic acid within the *ACes*. Also provided is a method of making one or more cell lines, comprising

- a) integrating into endogenous chromosomal DNA of a selected cell species, a multiplicity of heterologous recombination sites,
- 15 b) introducing a multiplicity of *ACes* under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites integrated within the cell's endogenous chromosomal DNA;
- c) isolating said multiplicity of *ACes*, thereby producing a library of  
20 *ACes* whereby a multiplicity of *ACes* have mutually exclusive portions of the endogenous chromosomal DNA therein;
- d) introducing the isolated multiplicity of *ACes* of step c) into a multiplicity of cells, thereby creating a library of cells;
- e) selecting different cells having mutually exclusive *ACes* therein  
25 and clonally expanding or differentiating said different cells into clonal cell cultures, thereby creating one or more cell lines.

These *ACes*, cell lines and methods utilize the site-specific recombination sites on platform *ACes* analogous YAC manipulation related to: the methods of generating terminal deletions in normal and

artificial chromosomes (e.g., *ACes*; as described in Vollrath et al., 1988, *PNAS, USA*, 85:6027-66031; and Pavan et al., *PNAS, USA*, 87:1300-1304); the methods of generating interstitial deletions in normal and artificial chromosomes (as described in Campbell et al., 1991, *PNAS, USA*, 888:5744-5748); and the methods of detecting homologous recombination between two *ACes* (as described in Cellini et al., 1991, *Nuc. Acid Res.*, 19(5):997-1000).

**5. Use of platform *ACes* in Pharmacogenomic/toxicology applications (development of "Reporter *ACes*")**

10 In addition to the placement of genes onto *ACes* chromosomes for therapeutic protein production or gene therapy, the platform can be engineered via the IntR lambda integrase to carry reporter-linked constructs (reporter genes) that monitor changes in cellular physiology as measured by the particular reporter gene (or a series of different reporter genes) readout.

15 The reporter linked constructs are designed to include a gene that can be detected (by for example fluorescence, drug resistance, immunohistochemistry, or transcript production, and the like) with well-known regulatory sequences that would control the expression of the detectable gene. Exemplary regulatory promoter sequences are well-known

20 in the art.

**A) Reporter *ACes* for drug pathway screening**

The *ACes* can be engineered to carry reporter-linked constructs that indicate a signal is being transduced through one or a number of pathways. For example, transcriptionally regulated promoters from genes at the end (or

25 any other chosen point) of particular signal transduction pathways could be engineered on the *ACes* to express the appropriate readout (either by fluorescent protein production or drug resistance) when the pathway is activated (or down-regulated as well). In one embodiment, a number of reporters from different pathways can be placed on an

ACes chromosome. Cells (and/or whole animals) containing such a Reporter ACes could be exposed to a variety of drugs or compounds and monitored for the effects of the drugs or compounds upon the selected pathway(s) by the reporter gene(s). Thus, drugs or compounds can be  
5 classified or identified by particular pathways they excite or down-regulate. Similarly, transcriptional profiles obtained from genomic array experiments can be biologically validated using the reporter ACes provided herein.

**B) Reporter ACes for toxic compound testing**

10 Environmental or man-made genotoxics can be tested in cell lines carrying a number of reporter-genes platform ACes linked to promoters that are transcriptionally regulated in response to DNA damage, induced apoptosis or necrosis, and cell-cycle perturbations. Furthermore, new drugs and/or compounds could be tested in a similar manner with the  
15 genotoxicant ACes reporter for their cellular/genetic toxicity by such a screen. Likewise, toxic compound testing could be carried out in whole transgenic animals carrying the ACes chromosome that measures genotoxicant exposure ("canary in a coal mine"). Thus, the same or similar type ACes could be used for toxicity testing in either a cell-based  
20 or whole animal setting. An example would include ACes that carry reporter-linked genes controlled by various cytochrome P450 profiled promoters and the like.

**C) Reporter ACes for individualized pharmacogenomics/drug profiling**

25 A common disease may arise via various mechanisms. In many instances there are multiple treatments available for a given disease. However, the success of a given treatment may depend upon the mechanism by which the disease originated and/or by the genetic background of the patient. In order to establish the most effective

treatment for a given patient one could utilize the *ACes* reporters provided herein. *ACes* reporters can be used in patient cell samples to determine an individualized drug regimen for the patient. In addition, potential polymorphisms affecting the transcriptional regulation of an individual's particular gene can be assessed by this approach.

**D) Reporter *ACes* for classification of similar patient tumors**

As with other diseases as described in 5.C) above, cancer cells arise via different mechanisms. Furthermore, as a cancerous cell propagates it may undergo genomic alterations. An *ACes* reporter transferred to cells of different patients having the same disease, i.e. similar cancers, could be used to categorize the particular cancer of each patient, thereby facilitating the identification of the most effective therapeutic regimen. Examples would include the validation of array profiling of certain classes of breast cancers. Subsequently, appropriate drug profiling could be carried out as described above.

**E) Reporter *ACes* as a "differentiation" sensor**

Using the *ACes* reporter as a "differentiation" sensor in stem cells or other progenitor cells in order to enrich by selection (either FACS based screening, drug selection and/or use of suicide gene) for a particular class of differentiated or undifferentiated cells. For example, in one embodiment, this assay could also be used for compound screening for small molecule modifiers of cell differentiation.

**F) Whole animal studies with Reporter *ACes***

Finally, with whole-body fluorescence imaging technology (Yang et al. (2000) PNAS 97:12278) any of the above Reporter *ACes* methods could be used in conjunction with whole-body imaging to monitor reporter genes within whole animals without sacrificing the animal. This would allow temporal and spatial analysis of expression patterns under a given set of conditions. The conditions tested may include for example, normal

differentiation of a stem cell, response to drug or compound treatment whether targeted to the diseased tissue or presented systemically, response to genotoxicants, and the like.

The following examples are included for illustrative purposes only  
5 and are not intended to limit the scope of the invention.

#### EXAMPLE 1

##### pFK161

Cosmid pFK161 (SEQ ID NO: 118) was obtained from Dr. Gyula Hadlaczky and contains a 9 kb *NotI* insert derived from a murine rDNA  
10 repeat (see clone 161 described in PCT Application Publication No. WO97/40183 by Hadlaczky *et al.* for a description of this cosmid). This cosmid, referred to as clone 161 contains sequence corresponding to nucleotides 10,232-15,000 in SEQ ID NO. 26. It was produced by inserting fragments of the megachromosome (see, U.S. Patent No.  
15 6,077,697 and International PCT application No. WO 97/40183). For example, H1D3, which was deposited at the European Collection of Animal Cell Culture (ECACC) under Accession No. 96040929, is a mouse-hamster hybrid cell line carrying this megachromosome into plasmid pWE15 (Stratagene, La Jolla, California; SEQ ID No. 31) as  
20 follows. Half of a 100  $\mu$ l low melting point agarose block (mega-plug) containing isolated SATACs was digested with *NotI* overnight at 37°C. Plasmid pWE15 was similarly digested with *NotI* overnight. The mega-plug was then melted and mixed with the digested plasmid, ligation buffer and T4 DNA ligase. Ligation was conducted at 16°C overnight. Bacterial  
25 DH5 $\alpha$  cells were transformed with the ligation product and transformed cells were plated onto LB/Amp plates. Fifteen to twenty colonies were grown on each plate for a total of 189 colonies. Plasmid DNA was isolated from colonies that survived growth on LB/Amp medium and analyzed by Southern blot hybridization for the presence of DNA that

hybridized to a pUC19 probe. This screening methodology assured that all clones, even clones lacking an insert but yet containing the pWE15 plasmid, would be detected.

Liquid cultures of all 189 transformants were used to generate  
5 cosmid minipreps for analysis of restriction sites within the insert DNA. Six of the original 189 cosmid clones contained an insert. These clones were designated as follows: 28 (~9-kb insert), 30 (~9-kb insert), 60 (~4-kb insert), 113 (~9-kb insert), 157 (~9-kb insert) and 161 (~9-kb insert). Restriction enzyme analysis indicated that three of the clones  
10 (113, 157 and 161) contained the same insert.

For sequence analysis the insert of cosmid clone no. 161 was subcloned as follows. To obtain the end fragments of the insert of clone no. 161, the clone was digested with *NotI* and *BamHI* and ligated with *NotI/BamHI*-digested pBluescript KS (Stratagene, La Jolla, California).  
15 Two fragments of the insert of clone no. 161 were obtained: a 0.2-kb and a 0.7-kb insert fragment. To subclone the internal fragment of the insert of clone no. 161, the same digest was ligated with *BamHI*-digested pUC19. Three fragments of the insert of clone no. 161 were obtained: a 0.6-kb, a 1.8-kb and a 4.8-kb insert fragment.

20 The insert corresponds to an internal section of the mouse ribosomal RNA gene (rDNA) repeat unit between positions 7551-15670 as set forth in GENBANK accession no. X82564, which is provided as SEQ ID NO. 18. The sequence data obtained for the insert of clone no. 161 is set forth in SEQ ID NOS. 19-25. Specifically, the individual  
25 subclones corresponded to the following positions in GENBANK accession no. X82564 (SEQ ID NO:18) and in SEQ ID NOS. 19-25:

Subclone	Start	End	Site	SEQ ID No.
	in X82564			
161k1	7579	7755	<i>NotI, BamHI</i>	19
161m5	7756	8494	<i>BamHI</i>	20
161m7	8495	10231	<i>BamHI</i>	21(shows only sequence corresponding to nt. 8495-8950), 22 (shows only sequence corresponding to nt. 9851- 10231)
161m12	10232	15000	<i>BamHI</i>	23 (shows only sequence corresponding to nt. 10232-10600), 24 (shows only sequence corresponding to nt. 14267-15000)
161k2	15001	15676	<i>NotI, BamHI</i>	25

The sequence set forth in SEQ ID NOs. 19-25 diverges in some positions from the sequence presented in positions 7551-15670 of GENBANK accession no. X82564. Such divergence may be attributable to random mutations between repeat units of rDNA.

For use herein, the rDNA insert from the clone was prepared by digesting the cosmid with *NotI* and *BglI* and was purified as described above. Growth and maintenance of bacterial stocks and purification of plasmids were performed using standard well known methods (see, *e.g.*, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press), and plasmids were purified from bacterial cultures using Midi - and Maxi-preps Kits (Qiagen, Mississauga, Ontario).

#### pDsRed1N1

This vector is available from Clontech (see SEQ ID No. 29) and encodes the red fluorescent protein (DsRed; Genbank accession no. AF272711; SEQ ID Nos. 39 and 40). DsRed, which has a vivid red fluorescence, was isolated from the IndoPacific sea anemone relative *Discosoma* species. The plasmid pDsRed1N1 (Clontech; SEQ ID No. 29) constitutively expresses a human codon-optimized variant of the



fluorescent protein under control of the CMV promoter. Unmodified, this vector expresses high levels of DsRed1 and includes sites for creating N-terminal fusions by cloning proteins of interest into the multiple cloning site (MCS). It is Kan and Neo resistant for selection in bacterial or  
5 eukaryotic cells.

#### **Plasmid pMG**

Plasmid pMG (InvivoGen, San Diego, California; see SEQ. ID. NO. 27 for the nucleotide sequence of pMG) contains the hygromycin phosphotransferase gene under the control of the immediate-early human  
10 cytomegalovirus (hCMV) enhancer/promoter with intron A. Vector pMG also contains two transcriptional units allowing for the coexpression of two heterologous genes from a single vector sequence.

The first transcriptional unit of pMG contains a multiple cloning site for insertion of a gene of interest, the hygromycin phosphotransferase  
15 gene (*hph*) and the immediate-early human cytomegalovirus (hCMV) enhancer/promoter with intron A (see, *e.g.*, Chapman *et al.* (1991) *Nuc. Acids Res.* 19:3979-3986) located upstream of *hph* and the multiple cloning site, which drives the expression of *hph* and any gene of interest inserted into the multiple cloning site as a polycistronic mRNA. The first  
20 transcriptional unit also contains a modified EMCV internal ribosomal entry site (IRES) upstream of the *hph* gene but downstream of the hCMV promoter and MCS for ribosomal entry in translation of the *hph* gene (see SEQ ID NO. 27, nucleotides 2736-3308). The IRES is modified by  
25 insertion of the constitutive *E. coli* promoter (EM7) within an intron (IM7) into the end of the IRES. In mammalian cells, the *E. coli* promoter is treated as an intron and is spliced out of the transcript. A polyadenylation signal from the bovine growth hormone (bGh) gene (see, *e.g.*, Goodwin and Rottman (1992) *J. Biol. Chem.* 267:16330-16334) and a pause site derived from the 3' flanking region of the human  $\alpha 2$

globin gene (see, *e.g.*, Enriquez-Harris *et al.* (1991) *EMBO J.* 10:1833-1842) are located at the end of the first transcription unit. Efficient polyadenylation is facilitated by inserting the flanking sequence of the bGh gene 3' to the standard AAUAAA hexanucleotide sequence.

5       The second transcriptional unit of pMG contains another multiple cloning site for insertion of a gene of interest and an EF-1 $\alpha$ /HTLV hybrid promoter located upstream of this multiple cloning site, which drives the expression of any gene of interest inserted into the multiple cloning site. The hybrid promoter is a modified human elongation factor-1 alpha (EF-1  
10   alpha) gene promoter (see, *e.g.*, Kim *et al.* (1990) *Gene* 91:217-223) that includes the R segment and part of the U5 sequence (R-U5') of the human T-cell leukemia virus (HTLV) type I long terminal repeat (see, *e.g.*, Takebe *et al.* (1988) *Mol. Cell. Biol.* 8:466-472). The Simian Virus 40 (SV40) late polyadenylation signal (see Carswell and Alwine (1989) *Mol.*  
15   *Cell. Biol.* 9:4248-4258) is located downstream of the multiple cloning site. Vector pMG contains a synthetic polyadenylation site for the first and second transcriptional units at the end of the transcriptional unit based on the rabbit  $\beta$ -globin gene and containing the AATAAA hexanucleotide sequence and a GT/T-rich sequence with 22-23  
20   nucleotides between them (see, *e.g.*, Levitt *et al.* (1989) *Genes Dev.* 3:1019-1025). A pause site derived from the C2 complement gene (see, Moreira *et al.* (1995) *EMBO J.* 14:3809-3819) is also located at the 3' end of the second transcriptional unit.

      Vector pMG also contains an ori sequence (ori pMB1) located  
25   between the SV40 polyadenylation signal and the synthetic polyadenylation site.

## EXAMPLE 2

**A. Construction of targeting vector and transfection into LMtk- cells for the generation of platform chromosomes**

A targeting vector derived from the vector pWE15 (GeneBank Accession # X65279) was modified by replacing the *SaI* (Klenow filled)/*SmaI* neomycin resistance containing fragment with the *PvuII*/*Bam*HI (Klenow filled) puromycin resistance containing fragment (isolated from plasmid pPUR, Clontech Laboratories, Inc. Palo Alto, CA; SEQ ID No. 30) resulting in plasmid pWEPuro. Subsequently a 9 Kb *NotI* fragment from the plasmid pFK161 (SEQ ID NO: 118) containing a portion of the mouse rDNA region was cloned into the *NotI* site of pWEPuro resulting in plasmid pWEPuro9K (Figure 2). The vector pWEPuro9K was digested with *SpeI* to linearize and transfected into LMtk- mouse cells. Puromycin resistant colonies were isolated and subsequently tested for artificial chromosome formation via fluorescent *in situ* hybridization (FISH) (using mouse major and minor DNA repeat sequences, the puromycin gene and telomeres sequences as probes), and fluorescent activated cell sorting (FACS). From this sort, a subclone was isolated containing an artificial chromosome, designated 5B11.12, which carries 4-8 copies of the puromycin resistance gene contained on the pWEPuro9K vector. FISH analysis of the 5B11.12 subclone demonstrated the presence of telomeres and mouse minor on the ACes. DOT PCR has been done on the 5B11.12 ACes revealing the absence of uncharacterized euchromatic regions on the ACes. A recombination site, such as an *att* or *loxP* engineering site or a plurality thereof, was introduced onto this ACes thereby providing a platform for site-specific introduction of heterologous nucleic acid.

**B. Targeting a single sequence specific recombination site onto platform chromosomes**

After the generation of the 5B11.12 platform, a single sequence-specific recombination site is placed onto the platform chromosome via homologous recombination. For this, DNA sequences containing the site-

specific recombination sequence can be flanked with DNA sequences of  
homology to the platform chromosome. For example, using the platform  
chromosome made from the pWEPuro9K vector, mouse rDNA sequences  
or mouse major satellite DNA can be used as homologous sequences to  
5 target onto the platform chromosome. A vector is designed to have these  
homologous sequences flanking the site-specific recombination site and,  
after the appropriate restriction enzyme digest to generate free ends of  
homology to the platform chromosome, the DNA is transfected into cells  
harboring the platform chromosome (Figure 3). Examples of site-specific  
10 cassettes that are targeted to the platform chromosome using either  
mouse rDNA or mouse major repeat DNA include the SV40-attP-hygro  
cassette and a red fluorescent protein (RFP) gene flanked by loxP sites  
(Cre/lox, see, *e.g.*, U.S. Patent No. 4,959,317 and description herein).  
After transfection and integration of the site-specific cassette,  
15 homologous recombination events onto the platform chromosome are  
subcloned and identified by FACS (*e.g.* screen and single cell subclone via  
expression of resistance or fluorescent marker) and PCR analysis.

For example, a vector can be constructed containing regions of the  
mouse rDNA locus flanking a gene cassette containing the SV40 early  
20 reporter-bacteriophage lambda attP site-hygromycin selectable marker  
(see Figure 4 and described below). The use of the bacteriophage lambda  
attP site for lambda integrase-mediated site-specific recombination is  
described below. Homologous recombination event of the SV40-attP-  
hygro cassette onto the platform chromosome was identified using PCR  
25 primers that detect the homologous recombination and further confirmed  
by FISH analysis. After identifying subcloned colonies containing the  
platform chromosome with a single site-specific recombination site, cells  
carrying the platform chromosome with a single site-specific

recombination site can now be engineered with site-specific recombinases (e.g. lambda INT, Cre) for integrating a target gene expression vector.

**C. Targeting a red fluorescent protein (RFP) gene flanked by loxP sites onto 5B11.12 platform**

As another example, while loxP recombination sites could have been introduced onto the *ACes* during *de novo* biosynthesis, it was thought that this might result in multiple segments of the *ACes* containing a high number of loxP sites, potentially leading to instability upon *Cre*-mediated recombination. A gene targeting approach was therefore devised to introduce a more limited number of loxP recombination sites into a locus of the 5B11-12 *ACes* containing introduced and possibly co-amplified endogenous rDNA sequences. Although there are more than 200 copies of rDNA genes in the haploid mouse genome distributed amongst 5-11 chromosomes (depending on strain), rDNA sequences were chosen as the target on the *ACes* since they represent a less frequent target than that of the satellite repeat sequences. Moreover, having observed much stronger pWEPuro9K hybridization to the 5B11-12 *ACes* than to other LMTK chromosomes and in light of the observation that the transcribed spacer sequences within the rDNA may be less conserved than the rRNA coding regions, it was contemplated that a targeting vector based on the rDNA gene segment in pWEPuro9K would have a higher probability of targeting to the *ACes* rather than to other LMTK chromosomes. Accordingly, a targeting vector, pBSFKLoxDsRedLox, was designed and constructed based on the rDNA sequences contained in pWEPuro9K.

The plasmid pBSFKLoxDsRedLox was generated in 4 steps. First, the *Not*I rDNA insert of pWEPuro9K (Figure 2) was inserted into pBS SK- (Stratagene) giving rise to pBSFK. Second, a loxP polylinker cassette was

generated by PCR amplification of pNEB193 (SEQ ID NO:32; New England Biolabs) using primers complementary to the M13 forward and reverse priming sites at their 3' end and a 34 bp 5' extension comprising a LoxP site. This cassette was reinserted into pNEB193 generating  
5 p193LoxMCSLox. Third, the DsRed gene from pDsRed1-N1 (SEQ ID NO:29; Clontech) was then cloned into the polylinker between the loxP sites generating p193LoxDsRedLox. Fourth, a fragment consisting of the DsRed gene flanked by loxP sites was cloned into a unique *NdeI* within the rDNA insert of pBSFK generating pBSFKLoxDsRedLox.

10 A gel purified 11 Kb *PmlI* / *EcoRV* fragment of pBSFKLoxDsRedLox was used for transfection. To detect targeted integration, PCR primers were designed from rDNA sequences within the 5' *NotI*-*PmlI* fragment of pWEPuro9K that is not present on the targeting fragment (5' primer) and sequence within the LoxDsRedLox cassette (3' primer). If the targeting  
15 DNA integrated correctly within the rDNA sequences, PCR amplification using these primers would give rise to a 2.3 Kb band. PCR reactions containing 1-4  $\mu$ l of genomic DNA were carried out according to the MasterTaq protocol (Eppendorf), using murine rDNA 5' primer (5'-CGGACAATGCGGTTGTGCGT-3'; SEQ ID NO:72) and DsRed 3' primer  
20 (5'-GGCCCCGTAATGCAGAAGAA-3'; SEQ ID NO:73) and PCR products were analyzed by agarose gel electrophoresis.

1.5X10<sup>6</sup> 5B11-12 LMTK<sup>-</sup> cells were transfected with 2  $\mu$ g of the pBSFKLoxDsRedLox targeting DNA described above using Lipofectamine Plus (Invitrogen). For flow sorting, harvested cells were suspended in  
25 medium and applied to the Becton Dickinson Vantage SE cell sorter, equipped with 488 nm lasers for excitation and 585/42 bandpass filter for optimum detection of RFP fluorescence. Cells were sorted using dPBS as sheath buffer. Negative control parental 5B11-12 cells and a positive control LMTK<sup>-</sup> cell line stably transfected with DsRed were used to

establish the selection gates. The RFP positive gated populations were recovered, diluted in medium supplemented with 1X penicillin-streptomycin (Invitrogen), then plated and cultured as previously described. After 4 rounds of enrichment, the percentage of RFP positive cells reached levels of 50% or higher. DNA from populations was analyzed by PCR for evidence of targeted integration. Ultimately, single cell subclones were established from positive pools and were analyzed by PCR and PCR-positive clones confirmed by FISH as described below. DNA was purified from pools or single cell clones using previously described methods set forth in Lahm et al., Transgenic Res., 1998; 7:131-134, or in some cases using a Wizard Genomic DNA purification kit (Promega). For FISH analysis, a biotinylated DsRed gene probe was generated by PCR using DsRed specific primers and biotin-labeled dUTP (5' RFP primer: 5'-GGTTTAAAGTGCGCTCCTCCAAGAACGTCATC-3', SEQ ID NO:74; and 3' RFP primer: 5'AGATCTAGAGCCGCGCTACAGGAACAGGTGGTGGCGGCC-3'; SEQ ID NO:75). To maximize the signal intensity of the DsRed probe, Tyramide amplification was carried out according to the manufacturers protocols (NEN).

The process of testing the feasibility of a more general targeting strategy that would not rely on enrichment *via* drug selection of stably transfected clones can be summarized as follows. A red fluorescent protein gene (RFP; encoded by the DsRed gene) was inserted between the loxP sites of the targeting vector to form pBSFKLoxDsRedLox. After transfection with pBSFKLoxDsRedLox, sequential rounds of high speed flow sorting and expansion of sorted cells in culture could then be used to enrich for stable transformants expressing RFP. In the event of targeted integration, PCR screening with primers that amplify from a spacer region within the segment of the 45s pre-rRNA gene in pWEPuro9K to a specific

anchor sequence within the DsRed gene in the targeting cassette would give rise to a diagnostic 2.3 Kb band. However, as rDNA clusters are found on several chromosomes, confirmation of targeting to an *ACes* would require fluorescence in situ hybridization (FISH) analysis. Finally, the flanking of the DsRed gene by loxP sites would allow for its removal and subsequent replacement with other genes of interest.

After transfection of the targeting sequence into 5B11-12 cells, enrichment for targeted clones was carried out using a combination of flow cytometry to detect red-fluorescing cells and PCR screening.

10 Ultimately 17 single cell subclones were identified as potential targeted clones by PCR and of these 16 were found by FISH to contain the DsRed integration event into the *ACes*. These subclones are referred to herein as D11-C4, D11-C12, D11-H3, C9-C9, C9-B9, C9-F4, C9-H8, C9-F2, C9-G8, C9-B6, C9-G3, C9-E12, C9-A11, C11-E3, C11-A9 and C11-H4. PCR

15 analysis of genomic DNA isolated from the D11-C4 subclone gave rise to a 2.3 Kb band, indicative of a targeted integration into an rDNA locus. Further analysis of the subclone by FISH analysis with a DsRed gene probe demonstrated integration of the LoxDsRedLox targeting cassette on the *ACes* co-localizing with one of the regions of rDNA staining seen on

20 the 5B11-12 *ACes*, consistent with a targeted integration into an rDNA locus of the *ACes*, while integrations on other chromosomes were not observed. Since transfected cells were maintained as heterogeneous populations through several cycles of sorting and replating it was not possible to estimate the frequency of targeted events. In most

25 mammalian cell lines the frequency of gene targeting via homologous recombination is roughly  $10^{-5}$ - $10^{-7}$  treated cells. Despite the low frequency of these events in mammalian cells, it is clear that an RFP expression based screening paradigm, coupled with PCR analysis, can effectively detect and enrich for such infrequent events in a large



population. In instances where drug selection is not possible or not desirable, such a system may provide a useful alternative. It was also verified that the modified *ACes* in subclone D11-C4 could be purified by flow cytometry. The results indicate that the flow karyogram of the D11-C4 subclone was unaltered from that of the 5B11-12 cell line. Thus, the D11-C4 *ACes* can be purified in high yield from native chromosomes of the host cell line.

**D. Reduction of LoxP on *ACes* to a single site.**

10 The strong hybridization signal detected by FISH on the *ACes* using the DsRed gene probe suggests that several copies of the targeting cassette may be present on the *ACes* in the D11-C4 line. This also suggests that multiple rDNA genes have been correctly targeted.

15 Accordingly, in certain embodiments where necessary, the number of loxP sites on the *ACes* can be reduced to a single site by *in situ* treatment with *Cre* recombinase, provided that the sites are co-linear. Such a process is described for multiple loxP-flanked integrations on a native mouse chromosome (Garrrick et al., Nature Genet., 1998, Jan;18(1):56-59). Reduction to a single loxP site on the D11-C4 *ACes* would result in the loss of the DsRed gene, forming the basis of a useful screen for this event.

25 For this purpose, a Cre expression plasmid pCX-Cre/GFP III has been generated by first deleting the EcoRI fragment of pCX-eGFP (SEQ ID NO:71) containing the eGFP coding sequence and replacing it with that of a PCR amplified Cre recombinase coding sequence (SEQ ID NO:58), generating pCX-Cre. Next, the Asel/Sspl fragment of pD2eGFP-N1 (containing the CMV promoter driving the D2EGFP gene with SV40 polyA signal; Clontech; SEQ ID NO:87) was inserted into the filled HindIII site of pCX-Cre, generating pCX-Cre\GFP III. Control plasmid pCX-CreRev\GFP

III was generated in similar fashion except that the Cre recombinase coding sequence was inserted in the antisense orientation. LMTK<sup>-</sup> cell line D11-C4 (containing first generation platform ACes with multiple loxP-DsRED sites) and 5B11-12 cell line (containing ACes with no loxP-DsRED sites) are maintained in culture as described above. D11C4 cells are transfected with 2  $\mu$ g of plasmid pCX-Cre\GFP III or 2  $\mu$ g pCX-CreRev\GFP III using Lipofectamine (Invitrogen) as previously described.

Forty-eight to seventy-two hours after transfection, transfected D11-C4 cells are harvested and GFP positive cells are sorted by cell cytometry using a FACSta Vantage cell sorter (Beckton-Dickinson) as follows: All D11-C4 cells transfected with pCX-Cre\GFP III or control plasmid pCX-CreRev\GFP III that exhibit GFP fluorescent higher than the gate level established by untransfected cells are collected and placed in culture a further 7-14 days. After 7-14 days the initial D11-C4 cells are harvested and analyzed by cell cytometry as follows: Untransfected D11-C4 cells are used to establish the gate that defines the RFP positive population, while 5B11-12 cells are used to set the RFP negative gate. The GFP positive population of D11-C4 transfected with pCX-Cre\GFP III should show decreased red fluorescence compared to pCX-CreRev\GFP III transfected or untransfected control D11-C4 cells. The cells exhibiting greatly decreased or no RFP expression are collected and single cell clones subsequently established. These clones will be expanded and analyzed by fluorescence *in-situ* hybridization and Southern blotting to confirm the removal of loxP-DsRed gene copies.

25

### EXAMPLE 3

**Construction of targeting vector and transfection into LMtk<sup>-</sup> cells for the generation of platform chromosomes containing multiple site-specific recombination sites**

An example of a selectable marker system for the creation of a chromosome-based platform is shown in Figure 4. This system includes a vector containing the SV40 early promoter immediately followed by (1) a 282 base pair (bp) sequence containing the bacteriophage lambda attP site and (2) the puromycin resistance marker. Initially a *PvuII/StuI* fragment containing the SV40 early promoter from plasmid pPUR (Clontech Laboratories, Inc., Palo Alto, CA; Seq ID No. 30) was subcloned into the *EcoRI/CRI* site of pNEB193 (a PUC19 derivative obtained from New England Biolabs, Beverly, MA; SEQ ID No. 32) generating the plasmid pSV40193. The only differences between pUC19 and pNEB193 are in the polylinker region. A unique *Ascl* site (GGCGCGCC) is located between the *BamHI* site and the *SmaI* site, a unique *PacI* site (TTAATTAA) is located between the *BamHI* site and the *XbaI* site and a unique *PmeI* site (GTTTAAAC) is located between the *PstI* site and the *SaI* site.

The attP site was PCR amplified from lambda genome (GenBank Accession # NC 001416) using the following primers:

attPUP: CCTTGCGCTAATGCTCTGTTACAGG SEQ ID No. 1

attPDWN: CAGAGGCAGGGAGTGGGACAAAATTG SEQ ID No. 2

After amplification and purification of the resulting fragment, the attP site was cloned into the *SmaI* site of pSV40193 and the orientation of the attP site was determined by DNA sequence analysis (plasmid pSV40193attP). The gene encoding puromycin resistance (Puro) was isolated by digesting the plasmid pPUR (Clontech Laboratories, Inc. Palo Alto, CA) with *AgeI/BamHI* followed by filling in the overhangs with Klenow and subsequently cloned into the *Ascl* site downstream of the attP site of pSV40193attP generating the plasmid pSV40193attPsensePUR (Figure 4; SEQ ID NO:113).

The plasmid pSV40193attPsensePUR was digested with *ScaI* and co-transfected with the plasmid pFK161 (SEQ ID NO: 118) into mouse LMtk- cells and platform artificial chromosomes were identified and isolated as described above. The process for generating this exemplary platform ACes containing multiple site-specific recombination sites is summarized in Figure 5. One platform ACes resulting from this experiment is designated B19-18. This platform ACes chromosome may subsequently be engineered to contain target gene expression nucleic acids using the lambda integrase mediated site-specific recombination system as described herein in Example 7 and 8.

#### EXAMPLE 4

##### **Lambda integrase mediated site-specific recombination of a RFP expressing vector onto artificial chromosomes**

In this example, a vector expressing the red fluorescent protein (RFP) was produced and recombined into the attP site residing on an artificial chromosome within LMTK- cells. This recombination is depicted in Figure 7.

##### **A. Construction of expression vectors containing wildtype and mutant lambda integrase**

Mutations at the glutamic acid at position 174 in the lambda integrase protein relaxes the requirement for the accessory protein IHF during recombination and DNA supercoiling *in vitro* (see, Miller *et al.* (1980) *Cell* 20:721-729; Lange-Gustafson *et al.* (1984) *J. Biol. Chem.* 259:12724-12732). Mutations at this site promote attP, attB intramolecular recombination in mammalian cells (Lorbach *et al.* (2000) *J. Mol. Biol* 296:1175-1181).

To construct nucleic acid encoding the mutant, lambda integrase was PCR amplified from bacteriophage lambda DNA (cl857 *ind Sam 7*; New England Biolabs) using the following primers:  
Lamint1 (SEQ ID No. 3)

TTCGAATTCATGGGAAGAAGGCGAAGTCATGAGCG)

Lamint2 (SEQ ID No. 4)

(TTCGAATTCTTATTTGATTTCAATTTGTCCCAC).

- 5 The resulting PCR product was digested with *EcoR* I and cloned into the *EcoR* I site of pUC19. Lambda integrase was mutated at amino acid position 174 using QuikChange Site-Directed Mutagenesis Kit (Stratagene) and the following oligos (generating a glutamic acid to arginine change at position 174):

LambdaINTE174R

- 10 (SEQ ID No. 6)

(CGCGCAGCAAAATCTAGAGTAAGGAGATCAAGACTTACGGCTGACG),

LamintR174rev (SEQ ID No. 7)

(CGTCAGCCGTAAGTCTTGATCTCCTTACTCTAGATTTTGCTGCGCG).

- 15 The resulting site directed mutant was confirmed by sequence analysis. The wildtype and mutant lambda genes were cloned into the *EcoR* I site of pCX creating pCX-LamInt (SEQ ID NO: 127) and pCXLamIntR (Figure 8; SEQ ID NO: 112).

- 20 The plasmid pCX (SEQ ID No. 70) was derived from plasmid pCXeGFP (SEQ ID No. 71). Excision of the *EcoRI* fragment containing the eGFP marker generated pCX. To generate plasmid pCXLamINTR (SEQ ID NO: 112) an *EcoRI* fragment containing the lambda integrase E174R (SEQ ID No. 37) mutation was cloned into the *EcoRI* site of pCX, and to generate plasmid pCX-LamINT, an *EcoRI* fragment containing the wild-type lambda integrase was cloned into the *EcoRI* site of pCX.

- 25 **B. Construction of integration vector containing attB and DsRed**

The plasmid pDsRedN1 (Clontech Laboratories, Palo Alto, CA; SEQ ID No. 29) was digested with *Hpa* I and ligated to the following annealed oligos:

attB1 (SEQ ID No. 8)

(TGAAGCCTGCTTTTTTATACTAACTTGAGCGAA)

attB2 (SEQ ID No. 9)

(TTCGCTCAAGTTAGTATAAAAAAGCAGGCTTCA)

The resulting vector (pDsRedN1-attB) was confirmed by PCR and  
5 sequence analysis.

**C. Transfection into LMtk- cells**

LM(tk-) cells containing the Prototype A ACes (L1-18; Chromos  
Molecular Systems Inc., Burnaby, BC Canada) were co-transfected with  
pDsRedN1 or pDsRedN1-attB and either pCXLamInt (SEQ ID NO: 127) or  
10 pCXLamIntR (SEQ ID NO: 112) using Lipofectamine Plus Reagent  
(LifeTechnologies, Gaithersburg, MD). The transfected cells were grown  
in DMEM (LifeTechnologies, Gaithersburg, MD) with 10% FBS (CanSera)  
and G418 (CalBiochem) at a concentration of 1 mg/ml.

**D. Enrichment by cell sorting**

15 The transfected cells were sorted using a FACs Vantage SE cell  
sorter (Becton Dickinson) to enrich for cells expressing DsRed. The cells  
were excited with a 488 nm Argon laser at 200 watts and cells  
fluorescing in the 585/42 detection channel were collected. The sorted  
cells were returned to growth medium for recovery and expansion. After  
20 three successive enrichments for cells expressing DsRed, single cell  
sorting into 96 well plates was performed using the same parameters.  
Duplicate plates of the single cell clones were made for PCR analysis.

**E. PCR analysis of single cell clones**

25 Pools of cells from each row and column of the 96 well plate were  
used for DNA isolation. DNA was prepared using a Wizard Genomic DNA  
purification kit (Promega Inc, Madison, WI). Nested PCR analysis on the  
DNA pools was performed to confirm the site-specific recombination  
event using the following primer sets:

attPdwn2 (SEQ ID No. 10)  
(TCTTCTCGGGCATAAGTCGGACACC)

CMVen (SEQ ID No. 11)  
(CTCACGGGGATTCCAAGTCTCCAC)

5 followed by:

attPdwn (SEQ ID No. 12)  
(CAGAGGCAGGGAGTGGGACAAAATTG)

CMVen2 (SEQ ID No. 13)  
(CAACTCCGCCCCATTGACGCAAATG).

10 The resulting PCR reactions were analyzed by gel electrophoresis and the potential individual clones containing the site-specific recombination event were identified by combining the PCR results of all of the pooled rows and columns for each 96 well plate. The individual clones were then further analyzed by PCR using the following primers that flank the  
15 recombination junction. L1for and F1rev flank the attR junction whereas REDfor and L2rev flank the attL junction (see Figure 7):

L1for (SEQ ID No. 14)  
AGTATCGCCGAACGATTAGCTCTTCA

20 F1rev (SEQ ID No. 15)  
GCCGATTTCGGCCTATTGGTTAAA

REDfor (SEQ ID No. 16)  
CCGCCGACATCCCCGACTACAAGAA

L2rev (SEQ ID No. 17)  
TTCCTTCGAAGGGGATCCGCCTACC.

25 **F. Sequence analysis of recombination junctions**

PCR products spanning the recombination junction were Topo-cloned into pcDNA3.1D/V5His (Invitrogen Inc., San Diego, CA) and then sequenced by cycle-sequencing. The clones were confirmed to have the correct *attR* and *attL* junctions by cycle sequencing.

30 **G. Fluorescent In Situ Hybridization (FISH)**

The cell lines containing the correct recombination junction sequence were further analyzed by fluorescent *in situ* hybridization (FISH)

by probing with the DsRed coding region labeled with biotin and visualizing with the Tyramide Signal Amplification system (TSA; NEN Life Science Products). The results indicate that the RFP sequence is present on the ACes.

5        **H. Southern analysis**

Genomic DNA was harvested from the cell lines containing an ACes with the correct recombinant event and digested with *EcoR* I. The digested DNAs were separated on a 0.7% agarose gel, transferred and fixed to a nylon membrane and probed with RFP coding sequences. The  
10 result showed that there is an integrated copy of RFP coding sequence in each clone.

**EXAMPLE 5**

**Delivery of a second gene encoding GFP onto the RFP platform ACes**

15        **A. Construction of integration vector containing attB and GFP (pD2eGFPIresPuroattB).**

The plasmid pIRESpuo2 (Clontech, Palo Alto, CA; SEQ ID NO: 88) was digested with *EcoR*I and *Not*I then ligated to the D2eGFP *EcoR*I-*Not*I fragment from pD2eGFP-N1 (Clontech, Palo Alto, CA) to create pD2eGFPIresPuro2. Subsequently, oligos encoding the attB site were  
20 annealed and ligated into the *Nru*I site of pD2eGFPIresPuro2 to create pD2eGFPIresPuroattB. The orientation of attB in the *Nru*I site was determined by PCR.

**B. Transfection of LMtk- cells**

The LMtk- cells containing the RFP platform ACes produced in  
25 Example 4, which has multiple attP sites, were co-transfected with pCXLamIntR and pD2eGFPIresPuroattB using LipofectAMINE PLUS reagent. Five  $\mu$ g of each vector was placed into a tube containing 750  $\mu$ l of DMEM (Dulbecco's modified Eagles Medium). Twenty  $\mu$ l of the Plus reagent was added to the DNA and incubated at room temperature for 15



minutes. A mixture of 30  $\mu$ l of lipofectamine and 750  $\mu$ l DMEM was added to the DNA mixture and incubated an additional 15 minutes at room temperature. The DNA mixture was then added dropwise to approximately 3 million cells attached to a 10cm dish in 5 mls of DMEM.

- 5 The cells were incubated 4 hours (37°C, 5% CO<sub>2</sub>) with the DNA-lipid mixture, after which DMEM with 20% fetal bovine serum was added to the dishes to bring the culture medium to 10% fetal bovine serum. The dishes were incubated at 37°C with 5% CO<sub>2</sub>.

- Plasmid pD2eGFPIresPuroattB has a puromycin gene
- 10 transcriptionally linked to the GFP gene *via* an IRES element. Two days after the transfection the cells were placed in medium containing puromycin at 4 $\mu$ g/ml to select for cells containing the pD2eGFPIresPuroattB plasmid integrated into the genome. Twenty-three clones were isolated after 17 days of selection with puromycin. These
- 15 clones were expanded and then analyzed for the presence of the GFP gene on the ACes by 2-color (RFP/biotin & GFP/digoxigenin) TSA-FISH (NEN) according to the manufacturers protocol. Sixteen of the 23 clones produced a positive FISH signal on the ACes with a GFP probe.

#### EXAMPLE 6

#### 20 Delivery Of ACes Into human Mesenchymal Stem Cells (hMSC)

##### A. Transfection

- Transfection conditions for the most efficient delivery of the ACes into hMSCs (Cambrex BioWhittaker Product Code PT-2501, lot# F0658, East Rutherford, New Jersey) were assayed using LipofectAMINE PLUS
- 25 and Superfect. One million prototype B ACes, which is a murine derived 60Mb ACes having primarily murine pericentric heterochromatin, and carrying a "payload" containing a hygromycin B selectable marker gene and a *lacZ* reporter gene (see , Telenius et al., 1999, Chrom. Res., 7:3-7; and Kereso et al., 1996, Chrom. Res., 4:226-239; each of which is

incorporated herein by reference in its entirety), were combined with 1-12  $\mu$ l of the transfection agent. In the case of LipofectAMINE PLUS, the PLUS reagent was combined with the ACes for 15 minutes followed by LipofectAMINE for a further 15 minutes. Superfect was complexed for 5 10 minutes at a ratio of 2  $\mu$ l Superfect per 1 million ACes. The ACes/transfection agent complex was then applied to 0.5 million recipient cells and the transfection was allowed to proceed according to the manufacturer's protocol. Percent transfected cells was determined on a FACS Vantage flow cytometer with argon laser tuned to 488 nm at 10 200mW and FITC fluorescence collected through a standard FITC 530/30 nm band pass filter. After 24 hours, IdUrd labeled ACes were delivered to human MSCs in the range of 30-50%, varying with transfection agent and dose. ACes delivery curves were generated from data collected in 15 experiments that varied the dose of the transfection reagents. Dose response curves of Superfect and LipofectAMINE PLUS, showing delivery of ACes into recipient hMSCs cells, were prepared, measured by transfer of IdUrd labeled ACes and detected by flow cytometry. Superfect shows maximum delivery in the range of 30-50% at doses greater than 2  $\mu$ l per million ACes. LipofectAMINE PLUS has a 42-48% delivery peak around 20 5-8  $\mu$ l per million ACes. These dose curves were then correlated with toxicity data to determine the transfection conditions that will allow for highest potential transfection efficiency. Toxicity was determined by a modified plating efficiency assay (de Jong et al., 2001, Chrom. Research, 9:475-485). The population's normalized plating efficiency (at maximum 25 % delivery doses) was in the range of 0.2 - 0.4 for Superfect and 0.5 - 0.6 with LipofectAMINE PLUS.

Due to the transfected population consisting of mixed cell types, flow cytometry allowed for the assessment of ACes delivery into each sub-population and the purification of the target population. Flow profiles

showing forward scatter (cell size) and side scatter (internal cell granularity) revealed three distinct hMSC populations that were gated into three regions: R3 (small cell region), R4 (medium cell region), R5 (large cell region). Transfection conditions were further optimized by re-analyzing delivery curves and assessing the differences in delivery to each sub-population. Dose response curves of Superfect and LipofectAMINE were prepared showing % delivery to each sub-population represented by the gating on basis of cell size and granularity properties of the mixed population. Three distinct hMSC populations were gated and % delivery dose curves generated. Using Superfect and LipofectAMINE PLUS the overall % delivery increased with cell size (80-90% delivery in large cells). LipofectAMINE PLUS at high doses (8-12  $\mu$ l per 1 million ACes) shows an increase in the overall proportion of chromosome transfer to the small population (10-20%). This suggests an advantage to using this transfection agent if the small-undifferentiated cell population is the desired target host cell.

**B. Expression from Genes on ACes IN hMSCs**

Following the delivery screening process conducted in section (A) above, the most promising results were subjected to further analyses to monitor expression and verify the presence of structurally intact ACes. The transfection conditions employed for these experiments were exactly the same as those that had been used during the screening process. Short-term expression was monitored by transfecting hMSCs with ACes containing a RFP gene (red fluorescent protein) set forth in Example 2C as "D11C4". The unselected population was harvested at 72-96 hours post transfection and % positive fluorescent cells measured by flow cytometry. RFP expression was in the range of 1-20%.

Long term-gene expression was assayed by selecting for hygromycin B resistant cells over a period of 7-10 days. Cytogenetic

analysis was done to detect presence of intact *ACes* by Fluorescent *In Situ* hybridization (FISH), where metaphase chromosomes were hybridized to a mouse major satellite-DNA probe (targeting murine pericentric heterochromatin) and a lambda probe (hybridizing to the *lacZ* gene). The human mesenchymal transfected culture could not undergo standard sub-cloning as diffuse colonies form with limited doublings available for expansion. Cytogenetic analysis was performed on the entire population, sampling over a period of 3-10 days post-transfection. The hygromycin resistant population was then blocked in mitosis with colchicine and analyzed for presence of intact *ACes* by FISH. Preliminary FISH results show approximately 2-8% of the hMSC-transfected population had an intact *ACes*. This compared to rat skeletal muscle myoblast clones, which were in the range of 60-95%. To increase the % of intact *ACes* in the hMSC-transfected population an enrichment step can be utilized as described in Example 2C.

#### C. Differentiation of The hMSCs

In initial experiments where transfected hMSCs cells have been induced to differentiate into adipose or osteocytes, the results indicate that the transfected cells appear to be differentiating at a rate comparable to the untransfected controls and the cultures are lineage specific as tested by microscopic examination, FISH, Oil Red O staining (adipocyte assay), and calcium secretion (osteocyte assay).

Accordingly, these results indicate that the artificial chromosomes (*ACes*) provided herein can be successfully transferred into hMSC target cells. Targeting MSCs (such as hMSCs) permits gene transfer into cells in an undifferentiated state where the cells are easier to expand and purify. The genetically modified cells can then be differentiated *in vitro* or injected into a site *in vivo* where the microenvironment will induce transformation into specific cell lineages.

#### EXAMPLE 7

##### **Delivery of a Promoterless Marker Gene to a Platform ACes**

Platform ACes containing pSV40attPsensePURO (Figure 4) were constructed as set forth in Examples 3 and 4.

5    **A.      Construction of Targeting Vectors.**

The base vector p18attBZeo (3166bp; SEQ ID NO: 114) was constructed by ligating the 1067bp *HindIII*-*SspI* fragment containing attBZeo, obtained from pLITattBZeo (SEQ ID NO:91), into pUC18 (SEQ ID NO: 122) digested with *HindIII* and *SspI*.

10        1. p18attBZEO-eGFP (6119bp; SEQ ID NO: 126) was constructed by inserting the 2977bp *SpeI*-*HindIII* fragment from pCXeGFP (SEQ ID NO:71; Okabe, *et al.* (1997) *FEBS Lett* 407:313-319) containing the eGFP gene into p18attBZeo (SEQ ID NO: 114) digested with *HindIII* and *XbaI*.

15        2. p18attBZEO-5'6XHS4eGFP (Figure 10; 7631bp; SEQ ID NO: 116) was constructed by ligating the 4465bp *HindIII* fragment from pCXeGFPattB(6XHS4)2 (SEQ ID NO: 123), which contains the eGFP gene under the regulation of the chicken beta actin promoter, 6 copies of the HS4 core element located 5' of the chicken beta actin promoter and the polyadenylation signal, into the *HindIII* site of p18attBZeo (SEQ ID NO: 20    114).

3. p18attBZEO-3'6XHS4eGFP (Figure 11; 7600bp; SEQ ID NO: 115) was created by removing the 5'6XHS4 element from p18attBZeo-(6XHS4)2eGFP (SEQ ID NO: 110). p18attBZeo-(6XHS4)2eGFP was digested with *EcoRV* and *SpeI*, treated with Klenow and religated to form 25    p18attBZeo3'6XHS4eGFP (SEQ ID NO: 115).

4. p18attBZEO-(6XHS4)2eGFP (Figure 12; 9080bp; SEQ ID NO: 110) was created in two steps. First, the *EcoRI*-*SpeI* fragment from pCXeGFPattB(6XHS4)2 (SEQ ID NO: 123), which contains 6 copies of the HS4 core element, was ligated into p18attBZeo (SEQ ID NO: 114)

digested with *EcoRI* and *XbaI* to create p18attBZeo6XHS4 (4615bp; SEQ ID NO: 117). Next, p18attBZeo6XHS4 was digested with *HindIII* and ligated to the 4465bp *HindIII* fragment from pCXeGFPattB(6XHS4)2 which contains the eGFP gene under the regulation of the chicken beta actin promoter, 6 copies of the HS4 core element located 5' of the chicken beta actin promoter and the polyadenylation signal.

Table 2

Targeting plasmid	No. zeocin resistant clones	No. clones with expected PCR product size	No. clones with correct sequence at recombination junction
p18attBZEO-eGFP	12	12	NT*
p18attBZEO-5'6XHS4eGFP	11	11	NT
p18attBZEO-3'6XHS4eGFP	11	11	NT
p18attBZEO-(6XHS4)2eGFP	9	9	4/4

\*NT = not tested

**B. Transfection and Selection with Drug.**

The mouse cell line containing the 2<sup>nd</sup> generation platform ACE, B19-38 (constructed as set forth in Example 3), was plated onto four 10cm dishes at approximately 5 million cells per dish. The cells were incubated overnight in DMEM with 10% fetal calf serum at 37°C and 5% CO<sub>2</sub>. The following day the cells were transfected with 5µg of each of the 4 vectors listed in Example 7.A. above and 5µg of pCXLamIntR (SEQ ID NO: 112), for a total of 10µg per 10cm dish. Lipofectamine Plus reagent was used to transfect the cells according to the manufacturers protocol. Two days post-transfection zeocin was added to the medium at 500µg/ml. The cells were maintained in selective medium until colonies formed. The colonies were then ring-cloned (see, e.g., McFarland, 2000, Methods Cell Sci, Mar;22(1):63-66).

**C. Analysis of Clones (PCR, SEQUENCING).**

Genomic DNA was isolated from each of the candidate clones with the Wizard kit (Promega) and following the manufacturers protocol. The following primer set was used to analyze the genomic DNA isolated from the zeocin resistant clones: 5PacSV40 –

- 5 CTGTTAATTAAGTGTGGAATGTGTG TCAGTTAGGGTG (SEQ ID NO:76);  
Antisense Zeo - TGAACAGGGTCACGTCGTCC (SEQ ID NO:77). PCR amplification with the above primers and genomic DNA from the site-specific integration of any of the 4 zeocin vectors would result in a 673bp PCR product.

- 10 As set forth in Table 2, of the 4 zeocin resistant candidate clones thusfar analyzed by PCR, all 4 exhibit the correct sequence for a site-specific integration event.

#### EXAMPLE 8

Integration of a PCR product by site-specific recombination.

- 15 In this example a gene is integrated onto the platform ACes by site-specific recombination without cloning said gene into a vector.

##### A. PCR PRIMER DESIGN.

- 20 PCR primers are designed to contain an attB site at the 5' end of one of the primers in the primer set. The remaining primers, which could be one or more than one primer, do not contain an attB site, but are complementary to sequences flanking the gene or genes of interest and any associated regulatory sequences. In first example, 2 primers (one containing an attB site) are used to amplify a selective gene such as puromycin.

- 25 In a second example as shown in Figure 13, the primer set includes primers 1 & 2 that amplify the GFP gene without amplification of an upstream promoter. Primer 1 contains the attB site at the 5' end of the oligo. Primers 3 & 4 are designed to amplify the IRES-blasticidin DNA sequences from the vector pIRESblasticidin. The 5'end of primer 3

contains sequences complementary to the 5' end of primer 2 such that annealing can occur between 5' ends of the two primers.

**B. PCR REACTION AND SUBSEQUENT LIGATION TO CREATE CIRCULAR MOLECULES FROM THE PCR PRODUCT**

5 In the first example set forth above in Section A, the two PCR primers are combined with a puromycin DNA template such as pPUR (Clontech), a heat stable DNA polymerase and appropriate conditions for DNA amplification. The resulting PCR product (attB-Puromycin) is then then purified and self-ligated to form a circular molecule.

10 In the second example set forth above in Section A, amplification of the GFP gene and IRES-blasticidin sequences is accomplished by combining primers 1 & 2 with DNA template pD2eGFP and primers 3 & 4 with template pIRESblasticidin under appropriate conditions to amplify the desired template. After initial amplification of the two products (attB-GFP  
15 & IRES-blasticidin) in separate reactions, a second round of amplification using both of the PCR products from the first round of amplification together with primers 1 and 4 amplifies the fusion product attB-GFP-IRES-blasticidin (Figure 13). This technique of using complementary sequences in primer design to create a fusion product is employed in *Saccharomyces*  
20 *cerevisiae* for allele replacement (Erdeniz *et al* (1997) *Gen Res* 7:1174-1183). The amplified product is then purified from the PCR reaction mixture by standard methods and ligated to form a circular molecule.

**C. INTRODUCTION OF PCR PRODUCT ONTO THE ACes USING A RECOMBINASE**

25 The circular PCR product is then be introduced to the platform ACes using the bacteriophage lambda integrase E174R. The introduction can be performed *in vivo* by transfecting the pCXLamIntR (SEQ ID NO: 112) vector encoding the lambda integrase mutant E174R together with the circularized PCR product into a cell line containing the platform ACE.



**D. SELECTION FOR MARKER GENE**

The marker gene (in this case either puromycin, blasticidin or GFP) is used to enrich the population for cells containing the proper integration event. A proper integration event in the second example (Figure 14) juxtaposes a promoter residing on the platform *ACes* 5' to the attB-GFP-IRES-Blasticidin PCR product, allowing for transcription of both GFP and blasticidin. If enrichment is done by drug selection, blasticidin is added to the medium on the transfected cells 24-48 hours post-transfection. Selection is maintained until colonies are formed on the plates. If enrichment is done by cell sorting, cells are sorted 2-4 days post-transfection to enrich for cells expressing the fluorescent marker (GFP in this case).

**E. ANALYSIS OF CLONES**

Clonal isolates are analyzed by PCR, FISH and sequence analysis to confirm proper integration events.

**EXAMPLE 9**

**Construction of a human platform *ACes* "ACE 0.1"**

**A. CONSTRUCTION OF THE TARGETING VECTOR pPACrDNA**

Genome Systems (IncyteGenomics) was supplied with the primers 5'HETS (GGGCCGAAACGATCTCAACCTATT; SEQ ID NO:78), and 3'HETS (CGCAGCGGCCCTCCTACTC; SEQ ID NO:79), which were used to amplify a 538bp PCR product homologous to nt 9680-10218 of the human rDNA sequences (GenBank Accession No. U13369) and used as a probe to screen a human genomic P1AC (P1 Artificial Chromosome) library constructed in the vector pCYPAC2 (Ioannou *et al.* (1994) *Nat. Genet.* 6(1): 84-89). Genome Systems clone #18720 was isolated in this screen and contains three repeats of human rDNA as assessed by restriction analysis. GS clone #18720, was digested with PmeI, a restriction enzyme unique to a single repeat of the human rDNA (45Kbp),

and then religated to form pPACrDNA (Figure 15). The insert in pPACrDNA was analyzed by restriction digests and sequence analysis of the 5' and 3' termini. The pPACrDNA, rDNA sequences are homologous to Genbank Accession #U13369, containing an insert of about 45 kB  
5 comprising a single repeat beginning from the end of one repeat at ~33980 (relative to the Genbank sequence) through the beginning of the next repeat up to approximately 35120 (the repeat offset from that listed in the GenBank file). Thus, the rDNA sequence is just over 1 copy of the repeat extending from 33980 (+/-10bp) to the end of the first repeat  
10 (43Kbp) and continuing into the second repeat to bp 35120 (+/-10bp).

**B. TRANSFECTION AND *ACes* FORMATION.**

Five hundred thousand MSU1.1 cells (Morgan et al., 1991, Exp. Cell Res., Nov;197(1):125-136; provided by Dr. Justin McCormick at Michigan State University) were plated per 6cm plate (3 plates total) and  
15 allowed to grow overnight. The cells were 70-80% confluent the following day. One plate was transfected with 15µg pPACrDNA (linearized with *Pme* I) and 2µg pSV40attPsensePuro (linearized with *Sca* I; see Example 3). The remaining plates were controls and were transfected with either 20µg pBS (Stratagene) or 20µg  
20 pSV40attBsensePuro (linearized with *Sca* I). All three plates were transfected using a CaPO<sub>4</sub> protocol.

**C. SELECTION OF PUROMYCIN RESISTANT COLONIES**

One day post-transfection the cells were "glycerol shocked" by the addition of PBS medium containing 10% glycerol for 30 seconds.  
25 Subsequently, the glycerol was removed and replaced with fresh DMEM. Four days post-transfection selective medium was added. Selective medium contains 1µg/ml puromycin. The transfection plates were maintained at 37°C with 5% CO<sub>2</sub> in selective medium for 2 weeks at which point colonies could be seen on the plate transfected with

pPACrDNA and pSV40attPsensePuro. The colonies were ring-cloned from the plate on day 17 post-selection and expanded in selective medium for analysis. Only two colonies (M2-2d & M2-2b) were able to proliferate in the selective medium after cloning. No colonies were seen on the control plates after 37 days in selective medium.

#### **D. ANALYSIS OF CLONES**

FISH analysis was performed on the candidate clones to detect *ACes* formation. Metaphase spreads from the candidate clones were probed in multiple probe combinations. In one experiment, the probes used were biotin-labeled human alphoid DNA (pPACrDNA) and digoxigenin-labeled mouse major DNA (pFK161) as a negative control. Candidate M2-2d was single cell subcloned by flow sorting and the candidate subclones were reanalyzed by FISH. Subclone 1B1 of M2-2d was determined to be a platform *ACes* and is also designated human Platform ACE 0.1.

#### **EXAMPLE 10**

##### **Site-specific integration of a marker gene onto a human platform ACE 0.1**

The promoterless delivery method was used to deliver a promoterless blasticidin marker gene onto the human platform *ACes* with excellent results. The human *ACes* platform with a promoterless blasticidin marker gene resulted in 21 of 38 blasticidin resistant clones displaying a PCR product of the expected size from the population co-transfected with pLIT38attBSRpolyA10 and pCXLamIntR (Figure 8; SEQ ID NOs. 111 and 112). Whereas, the population transfected with pBlueScript resulted in 0 blasticidin resistant colonies.

##### **A. CONSTRUCTION OF pLIT38attB-BSRpolyA10 & pLIT38attB-BSRpolyA2.**

The vector pLITMUS 38 (New England Biolabs; U.S. Patent No. 5,691,140; SEQ ID NO: 119) was digested with *EcoRV* and ligated to

two annealed oligomers, which form an attB site (attB1 5'-TGAAGCCTGCTTTTTTATACTAACTTGAGCGAA-3' (SEQ ID NO:8); attB2 5'- TTCGCTCAAGTTAGTATAAAAAAGCAGGCTTCA-3'; SEQ ID NO:9). This ligation reaction resulted in the vector pLIT38attB (SEQ ID NO: 120).

- 5 The blasticidin resistance gene and SV40 polyA site were PCR amplified with primers: 5BSD (ACCATGAAAACATTTAACATTTCTCAACA; SEQ ID NO:80) and SV40polyA (TTTATTTGTGAAATTTGTGATGCTATTGC; SEQ ID NO:81) using pPAC4 (Frengen, E., *et al.* (2000) Genomics 68 (2), 118-126; GenBank Accession No. U75992) as template. The blasticidin-
- 10 SV40polyA PCR product was then ligated into pLIT38attB at the *Bam*HI site, which was Klenow treated following digestion with *Bam*HI. pLIT38attB-BSDpolyA10 (SEQ ID NO: 111) and pLIT38attB-BSDpolyA2 (SEQ ID NO: 121) are the two resulting orientations of the PCR product ligated into the vector.

15 **B. TRANSFECTION OF MSU1.1 CELLS CONTAINING HUMAN PLATFORM ACE 0.1.**

- MSU1.1 cells containing human platform ACE 0.1 (see Example 9) were expanded and plated to five 10cm dishes with  $1.3 \times 10^6$  cells per dish. The cells were incubated overnight in DMEM with 10% fetal bovine
- 20 serum, at 37°C and 5% CO<sub>2</sub>. The following day the cells were transfected with 5µg of each plasmid as set forth in Table 3, for a total of 10µg of DNA per plate of cells transfected (see Table 3) using ExGen 500 *in vitro* transfection reagent (MBI fermentas, cat. no. R0511). The transfection was performed according to the manufacturers protocol.
- 25 Cells were incubated at 37°C with 5% CO<sub>2</sub> in DMEM with 10% fetal bovine serum following the transfection.

Table 3

Plate #	Plasmid 1	Plasmid 2	No. Bsd <sup>R</sup> Colonies
1	pBS	None	0
2	pCXLamInt	pLIT38attB-BSRpolyA10	16
3	pCXLamIntR	pLIT38attB-BSRpolyA10	40
4	pCXLamInt	pLIT38attB-BSRpolyA2	28
5	pCXLamIntR	pLIT38attB-BSRpolyA2	36

10 C. **SELECTION OF BLASTICIDIN RESISTANT CLONES.**

Three days following the transfection the cells were split from a 10 cm dish to two 15cm dishes. The cells were maintained in DMEM with 10% fetal bovine serum for 4 days in the 15 cm dishes. Seven days post-transfection blasticidin was introduced into the medium. Stably transfected cells were selected with 1 $\mu$ g/ml blasticidin. The number of colonies formed on each plate is listed in Table 3. These colonies were ring-cloned and expanded for PCR analysis. Upon expansion in blasticidin containing medium some clones failed to live and therefore do not have corresponding PCR data.

20 D. **PCR ANALYSIS**

Thirty-eight of the 40 clones from plate 3 grew after ring-cloning. Genomic DNA was isolated from these clones with the Promega Wizard Genomic cDNA purification kit, digested with *Eco*RI and used as template in a PCR reaction with the following primers: 3BSP – TTAATTTCGGG TATATTTGAGTGGA (SEQ ID NO:82); 5PacSV40 – CTGTTAATTAAGTGTGGAA TGTGTGTCAGTTAGGGTG (SEQ ID NO:76). The PCR conditions were as follows. 100ng of genomic DNA was

amplified with 0.5 $\mu$ l Herculanase polymerase (Stratagene) in a 50 $\mu$ l reaction that contained 12.5pmole of each primer, 2.5mM of each dNTP, and 1X Herculanase buffer (Stratagene). The reactions were placed in a PerkinElmer thermocycler programmed as follows: Initial denaturation at 95°C for 10 minutes; 35 cycles of 94°C for 1 minute, 53°C for 1 minute, 72°C for 1 minute, and 72°C for 1 minute; Final extension for 10 minutes at 72°C; and 4°C hold. If pLIT38attB-BSRpolyA10 integrates onto the human platform ACE 0.1 correctly, PCR amplification with the above primers should yield an 804bp product. Twenty-one of the 38 clones from plate 3 produced a PCR product of the expected 804bp size.

#### EXAMPLE 11

**Delivery of a Vector comprising a Promoterless Marker Gene and a gene encoding a therapeutic product to a Platform ACes**

Platform ACes containing pSV40attPsensePURO (Figure 4) were constructed as set forth in Examples 3 and 4.

##### A. CONSTRUCTION OF DELIVERY VECTORS

##### 1. Erythropoietin cDNA vector, p18EPOcDNA.

The erythropoietin cDNA was PCR amplified from a human cDNA library (E. Perkins *et al.*, 1999, *Proc. Natl. Acad. Sci. USA* 96(5): 2204-2209) using the following primers: EPO5XBA - TATCTAGAATGGGGGTGC ACGAATGTCCTGCC (SEQ ID NO: 83); EPO3BSI - TACGTACGTCATC TGTCCCCTGTCCTGCAGGC (SEQ ID NO: 84). The cDNA was amplified through two successive rounds of PCR using the following conditions: heat denaturation at 95°C for 3 minutes; 35 cycles of a 30 second denaturation (95°C), 30 seconds of annealing (60°C), and 1 minute extension (72°C); the last cycle is followed by a 7 minute extension at 72°C. BIO-X-ACT (BIOLINE) was used to amplify the erythropoietin cDNA from 2.5ng of the human cDNA library in the first round of amplification. Five  $\mu$ l of the first amplification product was used

as template for the second round of amplification. Two PCR products were produced from the second amplification with Taq polymerase (Eppendorf), each product was cloned into pCR2.1-Topo (Invitrogen) and sequenced. The larger PCR product contained the expected cDNA  
5 sequence for erythropoietin. The erythropoietin cDNA was moved from pTopoEPO into p18attBZeo(6XHS4)2eGFP (SEQ ID NO: 110). pTopoEPO was digested with BsiWI and XbaI to release a 588 bp EPO cDNA. BsrGI and BsiWI create compatible ends. The eGFP gene was removed from p18attBZeo(6XHS4)2eGFP by digestion with BsiWI and XbaI, the 8.3 Kbp  
10 vector backbone was gel purified and ligated to the 588 bp EPO cDNA to create p18EPOcDNA (SEQ ID NO: 124).

## **2. Genomic erythropoietin vector, p18genEPO.**

The erythropoietin genomic clone was PCR amplified from a human genomic library (Clontech) using the following primers: GENEPO3BSI -  
15 CGTACGTCATCTGTCCCCT GTCCTGCA (SEQ ID NO: 85); GENEPO 5XBA -TCTAGAATGGGGGT GCACGGTGAGTACT (SEQ ID NO: 86). The reaction conditions for the amplification were as follows: heat denaturation for 3 minutes (95°C); 30 cycles of a 30 second denaturation (95°C), 30 seconds annealing (from 65°C decreasing 0.5°C per cycle to  
20 50°C), and 3 minutes extension (72°C); 15 cycles of a 30 second denaturation (95°C), 30 seconds annealing (50°C), and 3 minute extension (72°C); the last cycle is followed by a 7 minute extension at 72°C. The erythropoietin genomic PCR product (2147 bp) was gel purified and cloned into pCR2.1Topo to create pTopogenEPO. Sequence  
25 analysis revealed 2bp substitutions and insertions in the intronic sequences of the genomic clone of erythropoietin. A partial digest with XbaI and complete digest with BsiWI excised the erythropoietin genomic insert from pTopogenEPO. The resulting 2158 bp genomic erythropoietin fragment was ligated into the 8.3 Kbp fragment resulting from the

digestion of p18attBZeo(6XHS4)2eGFP (SEQ ID NO: 110) with XbaI and BsrGI to create p18genEPO (SEQ ID NO: 125).

**B. TRANSFECTION AND SELECTION WITH DRUG**

5 The erythropoietin genomic and cDNA genes were each moved onto the platform ACes B19-38 (constructed as set forth in Example 3) by co-transfecting with pCXLamIntR. Control transfections were also performed using pCXLamInt (SEQ ID NO: 127) together with either p18EPOcDNA (SEQ ID NO: 124) or p18genEPO (SEQ ID NO: 125). Lipofectamine Plus was used to transfect the DNA's into B19-38 cells  
10 according to the manufacturer's protocol. The cells were placed in selective medium (DMEM with 10% FBS and Zeocin @ 500ug/ml) 48 hours post-transfection and maintained in selective medium for 13 days. Clones were isolated 15 days post-transfection.

**C. ANALYSIS OF CLONES (ELISA, PCR)**

15 **1. ELISA Assays**

Thirty clones were tested for erythropoietin production by an ELISA assay using a monoclonal anti-human erythropoietin antibody (R&D Systems, Catalogue # MAB287), a polyclonal anti-human erythropoietin antibody (R & D Systems, Catalogue # AB-286-NA) and alkaline  
20 phosphatase conjugated goat-anti-rabbit IgG (heavy and light chains) (Jackson ImmunoResearch Laboratories, Inc., Catalogue # 111-055-144). The negative control was a Zeocin resistant clone isolated from B19-38 cells transfected with p18attBZeo(6XHS4) (SEQ ID NO: 117; no insert control vector) and pCXLamIntR (SEQ ID NO: 112). The preliminary  
25 ELISA assay was executed as follows: 1) Nunc-Immuno Plates (MaxiSorb 96-well, Catalogue # 439454) were coated with 75µl of a 1/200 dilution (in Phosphate buffered Saline, pH 7.4 (PBS), Sigma Catalogue # P-3813) of monoclonal anti-human erythropoietin antibody overnight at 4°C. 2) The following day the plates were washed 3 times with 300µl PBS



containing 0.15% Tween 20 (Sigma, Catalogue # P-9416). 3) The plates were then blocked with 300 $\mu$ l of 1% Bovine Serum Albumin (BSA; Sigma Catalogue # A-7030) in PBS for 1 hour at 37°C. 4) Repeat the washes as in step 2. 5) The clonal supernatants (75 $\mu$ l per clone per well of 96-well plate) were then added to the plate and incubated for 1 hour at 37°C. The clonal supernatant analyzed in the ELISA assay had been maintained on the cells 7 days prior to analysis. 6) Repeat the washes of step 2. 7) Add 75 $\mu$ l of polyclonal anti-human erythropoietin antibody (1/250 dilution in dilution buffer (0.5% BSA, 0.01% Tween 20, 1X PBS, pH 7.4) and incubate 1 hour at 37°C. 8) Repeat washes of step 2. 9) Add 75 $\mu$ l of goat anti-rabbit conjugated alkaline phosphatase diluted 1/4000 in dilution buffer and incubate 1 hour at 37°C. 10) Repeat washes of step 2. 11) Add 75 $\mu$ l substrate, p-nitrophenyl phosphate (Sigma N2640), diluted to 1mg/ml in substrate buffer (0.1 Ethanolamine-HCl (Sigma, Catalogue # E-6133), 5mM MgCl<sub>2</sub> (Sigma, Catalogue # M-2393), pH 9.8). Incubate the plates in the dark for 1 hour at room temperature (22°C). 12) Read the absorption at 405nm (reference wavelength 495nm) on an Universal Microplate Reader (Bio-Tek Instruments, Inc., model # ELX800 UV). The erythropoietin standard curve was derived from readings of diluted human recombinant Erythropoietin (Roche, catalogue # 1-120-166; dilution range 125 - 7.8mUnits/ml). From this preliminary assay the 21 clones displaying the highest expression of erythropoietin were analyzed a second time in the same manner using medium supernatants that had been on the clones for 24 hours and a 1:3 dilution thereof.

## 25           2. PCR Analysis

Genomic DNA was isolated from the 21 clones with the best expression (as assessed by the initial ELISA assay above) as well as the B19-38 cell line and used for PCR analysis. Genomic DNA was isolated using the Wizard genomic DNA purification kit (Promega) according to the

manufacturers protocol. Amplification was performed on 100ng of genomic DNA as template with MasterTaq DNA Polymerase (Eppendorf) and the primer set 5PacSV40 – CTGTTAATTAAGTGTGGAATGTGTG TCAGTTAGGGTG (SEQ ID NO: 76) and Antisense Zeo -

- 5 TGAACAGGGTCACGTCGTCC (SEQ ID NO: 77). The amplification conditions were as follows: heat denaturation for 3 minutes (95°C); 30 cycles of a 30 second denaturation (95°C), 30 seconds annealing (from 65°C decreasing 0.5°C per cycle to 50°C), and 1 minutes extension (72°C); 15 cycles of a 30 second denaturation (95°C), 30 seconds  
10 annealing (50°C), and 1 minute extension (72°C); the last cycle is followed by a 10 minute extension at 72°C. PCR products were size separated by gel electrophoresis. Of the 21 clones analyzed 19 produced a PCR product of 650 bp as expected for a site-specific integration event. All nineteen clones were the result of transformations with p19EPOcDNA  
15 (5) or p18genEPO (14) and pCXLamIntR (i.e. mutant integrase). The remaining two clones, both of which were the result of transformation with p18genEPO (SEQ ID NO: 125) and pCXLamInt (i.e. wildtype integrase; SEQ ID NO: 127), produced a 400 bp PCR product.

#### EXAMPLE 12

#### 20 Preparation of a Transformation Vector Useful for the Induction of Plant Artificial Chromosome Formation

- Plant artificial chromosomes (PACs) can be generated by introducing nucleic acid, such as DNA, which can include a targeting DNA, for example rDNA or lambda DNA, into a plant cell, allowing the cell  
25 to grow, and then identifying from among the resulting cells those that include a chromosome with a structure that is distinct from that of any chromosome that existed in the cell prior to introduction of the nucleic acid. The structure of a PAC reflects amplification of chromosomal DNA, for example, segmented, repeat region-containing and heterochromatic

structures. It is also possible to select cells that contain structures that are precursors to PACs, for example, chromosomes containing more than one centromere and/or fragments thereof, and culture and/or manipulate them to ultimately generate a PAC within the cell.

5           In the method of generating PACs, the nucleic acid can be introduced into a variety of plant cells. The nucleic acid can include targeting DNA and/or a plant expressable DNA encoding one or multiple selectable markers (*e.g.*, DNA encoding bialophos (bar) resistance) or scorable markers (*e.g.*, DNA encoding GFP). Examples of targeting DNA  
10 include, but are not limited to, *N. tabacum* rDNA intergenic spacer sequence (IGS) and *Arabidopsis* rDNA such as the 18S, 5.8S, 26S rDNA and/or the intergenic spacer sequence. The DNA can be introduced using a variety of methods, including, but not limited to *Agrobacterium*-mediated methods, PEG-mediated DNA uptake and electroporation using,  
15 for example, standard procedures according to Hartmann *et al* [(1998) *Plant Molecular Biology* 36:741]. The cell into which such DNA is introduced can be grown under selective conditions and can initially be grown under non-selective conditions and then transferred to selective media. The cells or protoplasts can be placed on plates containing a  
20 selection agent to grow, for example, individual calli. Resistant calli can be scored for scorable marker expression. Metaphase spreads of resistant cultures can be prepared, and the metaphase chromosomes examined by FISH analysis using specific probes in order to detect amplification of regions of the chromosomes. Cells that have artificial chromosomes with  
25 functioning centromeres or artificial chromosomal intermediate structures, including, but not limited to, dicentric chromosomes, formerly dicentric chromosomes, minichromosomes, heterochromatin structures (*e.g.* sausage chromosomes), and stable self-replicating artificial chromosomal intermediates as described herein, are

identified and cultured. In particular, the cells containing self-replicating artificial chromosomes are identified.

The DNA introduced into a plant cell for the generation of PACs can be in any form, including in the form of a vector. An exemplary  
5 vector for use in methods of generating PACs can be prepared as follows.

For the production of artificial chromosomes, plant transformation vectors, as exemplified by pAgIIa and pAgIIb, containing a selectable marker, a targeting sequence, and a scorable marker were constructed using procedures well known in the art to combine the various fragments.  
10 The vectors can be prepared using vector pAg1 as a base vector and inserting the following DNA fragments into pAg1: DNA encoding  $\beta$ -glucoronidase under the control of the nopaline synthase (NOS) promoter fragment and flanked at the 3' end by the NOS terminator fragment, a fragment of mouse satellite DNA and an *N. tabacum* rDNA intergenic  
15 spacer sequence (IGS). In constructing plant transformation vectors, vector pAg2 can also be used as the base vector.

#### 1. Construction of pAg1

Vector pAg1 (SEQ. ID. NO: 89) is a derivative of the CAMBIA vector named pCambia 3300 (Center for the Application of Molecular  
20 Biology to International Agriculture, i.e., CAMBIA, Canberra, Australia; [www.cambia.org](http://www.cambia.org)), which is a modified version of vector pCambia 1300 to which has been added DNA from the bar gene conferring resistance to phosphinothricin. The nucleotide sequence of pCambia 3300 is provided in SEQ. ID. NO: 90. pCambia 3300 also contains a lacZ alpha sequence  
25 containing a polylinker region.

pAg1 was constructed by inserting two new functional DNA fragments into the polylinker of pCambia 3300: one sequence containing an attB site and a promoterless zeomycin resistance-encoding DNA flanked at the 3' end by a SV40 polyA signal sequence, and a second

sequence containing DNA from the hygromycin resistance gene (hygromycin phosphotransferase) conferring resistance to hygromycin for selection in plants. Although the zeomycin-SV40 polyA signal fusion is not expected to function in plant cells, it can be activated in mammalian  
5 cells by insertion of a functional promoter element into the attB site by site-specific recombination catalyzed by the Lambda att integrase. Thus, the inclusion of the attB-zeomycin sequences allows for evaluation of functionality of plant artificial chromosomes in mammalian cells by activation of the zeomycin resistance-encoding DNA, and provides an att  
10 site for further insertion of new DNA sequences into plant artificial chromosomes formed as a result of using pAg1 for plant transformation. The second functional DNA fragment allows for selection of plant cells with hygromycin. Thus, pAg1 contains DNA from the bar gene conferring resistance to phosphinothricin, DNA from the hygromycin resistance gene,  
15 both resistance-encoding DNAs under the control of a separate cauliflower mosaic virus (CaMV) 35S promoter, and the attB-promoterless zeomycin resistance-encoding DNA.

pAg1 is a binary vector containing *Agrobacterium* right and left T-DNA border sequences for use in *Agrobacterium*-mediated transformation  
20 of plant cells or protoplasts with the DNA located between the border sequences. pAg1 also contains the pBR322 Ori for replication in *E.coli*. pAg1 was constructed by ligating *HindIII/PstI*-digested p3300attBZeo with *HindIII/PstI*-digested pBSCaMV35SHyg as follows.

**a. Generation of p3300attBZeo**

25 Plasmid pCambia 3300 was digested with *PstI/Ec*136 II and ligated with *PstI/StuI*-digested pLITattBZeo (the nucleotide sequence of pLITattBZeo is provided in SEQ. ID. NO: 91), which contains DNA encoding the zeocin resistance gene and an attB Integrase recognition sequence, to generate p3300attBZeo which contains an attB site, a  
30 promoterless

zeomycin resistance-encoding DNA flanked at the 3' end by a SV40 polyA signal, and a reconstructed *Pst*I site.

**b. Generation of pBSCaMV35SHyg**

A DNA fragment containing DNA encoding hygromycin

- 5 phosphotransferase flanked by the CaMV 35S promoter and the CaMV 35S polyA signal sequence was obtained by PCR amplification of plasmid pCambia 1302 (GenBank Accession No. AF234298 and SEQ. ID. NO: 92). The primers used in the amplification reaction were as follows:

CaMV35SpolyA:

- 10 5'-CTGAATTAACGCCGAATTAATTCGGGGGATCTG-3' SEQ. ID. NO: 93

CaMV35Spr:

5'-CTAGAGCAGCTTGCCAACATGGTGGAGCA-3' SEQ. ID. NO: 94

The 2100-bp PCR fragment was ligated with *Eco*RV-digested pBluescript II SK+ (Stratagene, La Jolla, CA, U.S.A.) to generate pBSCaMV35SHyg.

- 15 **c. Generation of pAg1**

To generate pAg1, pBSCaMV35SHyg was digested with *Hind*III/*Pst*I and ligated with *Hind*III/*Pst*I-digested p3300attBZeo. Thus, pAg1 contains the pCambia 3300 backbone with DNA conferring resistance to phosphinothricin and hygromycin under the control of

- 20 separate CaMV 35S promoters, an attB-promoterless zeomycin resistance-encoding DNA recombination cassette and unique sites for adding additional markers, *e.g.*, DNA encoding GFP. The attB site can be used as described herein for the addition of new DNA sequences to plant artificial chromosomes, including PACs formed as a result of using the
- 25 pAg1 vector, or derivatives thereof, in the production of PACs. The attB site provides a convenient site for recombinase-mediated insertion of DNAs containing a homologous att site.

**2. pAG2**

The vector pAg2 (SEQ. ID. NO: 95) is a derivative of vector pAg1 formed by adding DNA encoding a green fluorescent protein (GFP), under the control of a NOS promoter and flanked at the 3' end by a NOS polyA signal, to pAg1. pAg2 was constructed as follows. A DNA fragment  
5 containing the NOS promoter was obtained by digestion of pGEM-T-NOS, or pGEMEasyNOS (SEQ. ID. NO: 96), containing the NOS promoter in the cloning vector pGEM-T-Easy (Promega Biotech, Madison, WI, U.S.A.), with *Xba*I/*Nco*I and was ligated to a *Xba*I/*Nco*I fragment of pCambia 1302 containing DNA encoding GFP (without the CaMV 35S promoter) to  
10 generate p1302NOS (SEQ. ID. NO: 97) containing GFP-encoding DNA in operable association with the NOS promoter. Plasmid p1302NOS was digested with *Sma*I/*Bsi*WI to yield a fragment containing the NOS promoter and GFP-encoding DNA. The fragment was ligated with *Pme*I/*Bsi*WI-digested pAg1 to generate pAg2. Thus, pAg2 contains DNA  
15 from the bar gene conferring resistance to phosphinothricin, DNA conferring resistance to hygromycin, both resistance-encoding DNAs under the control of a cauliflower mosaic virus 35S promoter, DNA encoding kanamycin resistance, a GFP gene under the control of a NOS promoter and the attB-zeomycin resistance-encoding DNA. One of skill in  
20 the art will appreciate that other fragments can be used to generate the pAg1 and pAg2 derivatives and that other heterologous DNA can be incorporated into pAg1 and pAg2 derivatives using methods well known in the art.

### 3. pAgIIa and pAgIIb transformation vectors

25 Vectors pAgIIa and pAgIIb were constructed by inserting the following DNA fragments into pAg1: DNA encoding  $\beta$ -glucuronidase, the nopaline synthase terminator fragment, the nopaline synthase (NOS) promoter fragment, a fragment of mouse satellite DNA and an *N. tabacum*

rDNA intergenic spacer sequence (IGS). The construction of pAgIIa and pAgIIb was as follows.

An *N. tabacum* rDNA intergenic spacer (IGS) sequence (SEQ. ID. NO: 98; see also GenBank Accession No. YO8422; see also Borysyuk *et al.* (2000) *Nature Biotechnology* 18:1303-1306; Borysyuk *et al.* (1997) *Plant Mol. Biol.* 35:655-660; U.S. Patent Nos. 6,100,092 and 6,355,860) was obtained by PCR amplification of tobacco genomic DNA. The IGS can be used as a targeting sequence by virtue of its homology to tobacco rDNA genes; the sequence is also an amplification promoter sequence in plants. This fragment was amplified using standard PCR conditions (*e.g.*, as described by Promega Biotech, Madison, WI, U.S.A.) from tobacco genomic DNA using the primers shown below:

NTIGS-FI

5'- GTG CTA GCC AAT GTT TAA CAA GAT G- 3' (SEQ ID No. 99) and

15 NTIGS-RI

5'-ATG TCT TAA AAA AAA AAA CCC AAG TGA C- 3' (SEQ ID No. 100)

Following amplification, the fragment was cloned into pGEM-T Easy to give pIGS-I. A fragment of mouse satellite DNA (Msat1 fragment; GenBank Accession No. V00846; and SEQ ID No. 101) was amplified via PCR from pSAT-1 using the following primers:

MSAT-F1

5'- AAT ACC GCG GAA GCT TGA CCT GGA ATA TCG C -3'(SEQ ID No. 102) and

MSAT-RI

25 5'-ATA ACC GCG GAG TCC TTC AGT GTG CA T- 3' (SEQ ID No. 103)

This amplification added a *SacII* and a *HindIII* site at the 5' end and a *SacII* site at the 3' end of the PCR fragment. This fragment was then cloned into the *SacII* site in pIGS-1 to give pMIGS-1, providing a eukaryotic



centromere-specific DNA and a convenient DNA sequence for detection via FISH.

A functional marker gene containing a NOS-promoter:GUS:NOS terminator fusion was then constructed containing the NOS promoter  
5 (GenBank Accession No. U09365; SEQ ID No. 104), *E. coli*  
 $\beta$ -glucuronidase coding sequence (from the GUS gene; GenBank  
Accession No. S69414; and SEQ ID No. 105), and the nopaline synthase  
terminator sequence (GenBank Accession No. U09365; SEQ ID No. 107).  
The NOS promoter in pGEM-T-NOS was added to a promoterless GUS  
10 gene in pBlueScript (Stratagene, La Jolla, CA, U.S.A.) using *NotI/Spel* to  
form pNGN-1, which has the NOS promoter in the opposite orientation  
relative to the GUS gene.

pMIGS-1 was digested with *NotI/Spel* to yield a fragment  
containing the mouse major satellite DNA and the tobacco IGS which was  
15 then added to *NotI*-digested pNGN-1 to yield pNGN-2. The NOS promoter  
was then re-oriented to provide a functional GUS gene, yielding pNGN-3,  
by digestion and religation with *Spel*. Plasmid pNGN-3 was then digested  
with *HindIII*, and the *HindIII* fragment containing the  $\beta$ -glucuronidase  
coding sequence and the rDNA intergenic spacer, along with the Msat  
20 sequence, was added to pAG-1 to form pAgIIa (SEQ ID NO: 108), using  
the unique *HindIII* site in pAg1 located near the right T-DNA border of  
pAg1, within the T-DNA region.

Another plasmid vector, referred to as pAgIIb, was also recovered,  
which contained the inserted *HindIII* fragment (SEQ ID NO: 108) in the  
25 opposite orientation relative to that observed in pAgIIa. Thus, pAgIIa and  
pAgIIb differ only in the orientation of the *HindIII* fragment containing the  
mouse major satellite sequence, the GUS DNA sequence and the IGS  
sequence. The nucleotide sequence of pAgIIa is provided in SEQ. ID. NO:  
109.

**-138-**

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

**WHAT IS CLAIMED IS:**

1. A eukaryotic chromosome comprising one or a plurality of *att* site(s), wherein:
  - an *att* site is heterologous to the chromosome; and
  - 5 an *att* site permits site-directed integration in the presence of lambda integrase.
2. The eukaryotic chromosome of claim 1, wherein the *att* sites are selected from the group consisting of *attP* and *attB* or *attL* and *attR*, or variants thereof.
- 10 3. The eukaryotic chromosome of claim 1 that is an artificial chromosome.
4. The eukaryotic chromosome of claim 1 that is an artificial chromosome expression system (ACes).
5. The eukaryotic chromosome of claim 4 that is predominantly  
15 heterochromatin.
6. The chromosome of claim 1 that is an artificial chromosome that contains no more than about 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.
7. The chromosome of claim 1 that is a plant chromosome.
- 20 8. The chromosome of claim 1 that is an animal chromosome.
9. The chromosome of claim 7 that is a plant artificial chromosome.
10. The chromosome of claim 8 that is an animal artificial chromosome.
- 25 11. The chromosome of claim 8 that is a mammalian chromosome.
12. The chromosome of claim 11 that is a mammalian artificial chromosome.

13. The chromosome of claim 6 that is an artificial chromosome expression system (*ACes*).
14. A platform artificial chromosome expression system (*ACes*) comprising one or a plurality of sites that participate in recombinase catalyzed recombination.
15. The *ACes* of claim 14 that contains one site.
16. The *ACes* of claim 14 that is predominantly heterochromatin.
17. The *ACes* of claim 14 that contains no more than about 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% euchromatin.
- 10 18. The *ACes* of claim 14 that is a plant *ACes*.
19. The *ACes* of claim 14 that is an animal *ACes*.
20. The *ACes* of claim 14 that is selected from a fish, insect, reptile, amphibian, arachnid or a mammalian *ACes*.
21. The *ACes* of claim 14 that is a fish *ACes*.
- 15 22. The artificial chromosome expression system (*ACes*) of claim 14, wherein the recombinase and site(s) are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase system of phage Mu, the Cin recombinase system, the Pin recombinase system of *E. coli* and the R/RS system of the pSR1 plasmid, or any combination thereof.
- 20 23. A method of introducing heterologous nucleic acid into a chromosome, comprising:
- contacting a chromosome of any of claims 1 or 14 with a nucleic acid molecule comprising both the heterologous nucleic acid and a
- 25 recombination site, in the presence of a recombinase that promotes recombination between the sites in the chromosome and in the nucleic acid molecule.

24. The method of claim 23, wherein the recombinase is selected from the group consisting of Cre, Gin, Cin, Pin, FLP, a phage integrase and R from the pSR1 plasmid.

25. The method of claim 23, wherein the nucleic acid molecule  
5 encodes a therapeutic protein, antisense nucleic acid, or comprises an artificial chromosome.

26. The method of claim 25, wherein the nucleic acid molecule comprises a yeast artificial chromosomes (YAC), a bacterial artificial chromosome (BAC) or an insect artificial chromosome (IAC).

10 27. A combination, comprising, the chromosome of claim 1 and a first vector comprising the cognate recombination site, wherein the cognate recombination site is a site that recombines with the site engineered into the chromosome.

28. The combination of claim 27, further comprising nucleic acid  
15 encoding a recombinase, wherein the nucleic acid is on a second vector or on the first vector, or on the ACes under an inducible promoter.

29. The combination of claim 28, wherein the recombinase and sites are from the Cre/lox system of bacteriophage P1, the int/att system of lambda phage, the FLP/FRT system of yeast, the Gin/gix recombinase  
20 system of phage Mu, the Pin recombinase system of *E. coli* and the R/RS system of the pSR1 plasmid, or any combination thereof.

30. The combination of claim 28, wherein a vector is the plasmid pCXLamIntR.

31. The combination of claim 27, wherein a vector is the plasmid  
25 pDsRedN1-attB.

32. A kit, comprising the combination of claim 27 and optionally instructions for introducing heterologous nucleic acid into the chromosome.

33. A method for introducing heterologous nucleic acid into a platform artificial chromosome, comprising:

(a) mixing an artificial chromosome comprising at least a first recombination site and a vector comprising at least a second  
5 recombination site and the heterologous nucleic acid;

(b) incubating the resulting mixture in the presence of at least one recombination protein under conditions whereby recombination between the first and second recombination sites is effected, thereby introducing the heterologous nucleic acid into the artificial chromosome.

10 34. The method of claim 33, wherein the artificial chromosome is an *ACes*.

35. The method of claim 33, wherein said mixing step (a) is conducted in cells *ex vivo*.

15 36. The method of claim 33, wherein said mixing step (a) is conducted extracellularly in an *in vitro* reaction mixture.

37. The method of claim 33, wherein the at least one recombination protein is encoded by a bacteriophage selected from the group consisting of bacteriophage lambda, phi 80, P22, P2, 186, P4 and P1.

20 38. The method of claim 37, wherein the at least one recombination protein is encoded by bacteriophage lambda, or mutants thereof.

39. The method of claim 33, wherein at least one recombination protein is selected from the group consisting of *Int*, *IHF*, *Xis* and *Cre*,  $\gamma\delta$ ,  
25 *Tn3* resolvase, *Hin*, *Gin*, *Cin* and *Flp*.

40. The method of claim 32, wherein the recombination sites are selected from the group consisting of *att* and *lox P* sites.

41. The method of claim 33, wherein the first and/or second recombination site contains at least one mutation that removes one or more stop codons.

42. The method of claim 33, wherein the first and/or second recombination site contains at least one mutation that avoids hairpin formation.

43. The method of claim 33, wherein the first and/or second recombination site comprises at least a first nucleic acid sequence selected from the group consisting of SEQ ID NOs:41-56:

- 10 a) RKYCWGCTTTYKTRTACNAASTSGB (m-att) (SEQ ID NO:41);
- b) AGCCWGCTTTYKTRTACNAACTSGB (m-attB) (SEQ ID NO:42);
- c) GTTCAGCTTTCCKTRTACNAACTSGB (m-attR) (SEQ ID NO:43);
- d) AGCCWGCTTTCCKTRTACNAAGTSGB (m-attL) (SEQ ID NO:44);
- e) GTTCAGCTTTYKTRTACNAAGTSGB (m-attP1) (SEQ ID NO:45);
- 15 f) AGCCTGCTTTTTTGTACAAACTTGT (attB1) (SEQ ID NO:46);
- g) AGCCTGCTTCTTGTACAAACTTGT (attB2) (SEQ ID NO:47);
- h) ACCCAGCTTCTTGTACAAACTTGT (attB3) (SEQ ID NO:48);
- i) GTTCAGCTTTTTTGTACAAACTTGT (attR1) (SEQ ID NO:49);
- j) GTTCAGCTTCTTGTACAAACTTGT (attR2) (SEQ ID NO:50);
- 20 k) GTTCAGCTTCTTGTACAAAGTTGG (attR3) (SEQ ID NO:51);
- l) AGCCTGCTTTTTTGTACAAAGTTGG (attL1) (SEQ ID NO:52);
- m) AGCCTGCTTCTTGTACAAAGTTGG (attL2) (SEQ ID NO:53);
- n) ACCCAGCTTCTTGTACAAAGTTGG (attL3) (SEQ ID NO:54);
- o) GTTCAGCTTTTTTGTACAAAGTTGG (attP1) (SEQ ID NO:55);
- 25 p) GTTCAGCTTCTTGTACAAAGTTGG (attP2, P3) (SEQ ID NO: 56);

and a corresponding or complementary DNA or RNA sequence, wherein R=A or G, K=G or T/U, Y=C or T/U, W=A or T/U, N=A or C or G or T/U, S=C or G, and B=C or G or T/U; and

the core region does not contain a stop codon in one or more reading frames.

44. The method of claim 33, wherein the first and/or second recombination site comprises at least a first nucleic acid sequence  
5 selected from the group consisting of a mutated att recombination site containing at least one mutation that enhances recombinational specificity, a complementary DNA sequence thereto, and an RNA sequence corresponding thereto.

45. The method of claim 33, wherein the vector comprising the  
10 second site further encodes at least one selectable marker.

46. The method of claim 45, wherein the marker is a promoterless marker, which, upon recombination is under the control of a promoter and is thereby expressed.

47. The method of claim 46, wherein the first recombination site  
15 is attP and is in the sense orientation prior to recombination.

48. The method of claim 46, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic, fluorescent, or capable of being bound by an antibody and FACs sorted.

20 49. The method of claim 48, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase (hisD).

50. A cell comprising, the chromosome of claim 1.

25 51. The cell of claim 50, wherein the cell is a nuclear donor cell.

52. The cell of claim 50, wherein the cell is a stem cell.

53. The stem cell of claim 52, wherein said stem cell is human and is selected from the group consisting of a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell and an embryonic stem cell.



54. The cell of claim 50, wherein the cell is mammalian.

55. The cell of claim 54, wherein the mammal is selected from the group consisting of humans, primates, cattle, pigs, rabbits, goats, sheep, mice, rats, guinea pigs, hamsters, cats, dogs, and horses.

5 56. The cell of claim 50, wherein the cell is a plant cell.

57. A cell comprising the platform ACes of claim 14.

58. The cell of claim 57, wherein the cell is a nuclear donor cell.

59. The cell of claim 57, wherein the cell is a stem cell.

60. The stem cell of claim 59, wherein said stem cell is human  
10 and is selected from the group consisting of a mesenchymal stem cell, a hematopoietic stem cell, an adult stem cell and an embryonic stem cell.

61. A human mesenchymal cell comprising an artificial  
chromosome.

62. The human mesenchymal cell of claim 61, wherein said  
15 artificial chromosome is an ACes.

63. The human mesenchymal cell of claim 62, wherein the ACes  
is a platform-ACes.

64. A method for introducing heterologous nucleic acid into the  
mesenchymal cell of claim 63, comprising:

20 (a) introducing into the cell of claim 63, wherein the platform-ACes  
has a first recombination site, a vector comprising at least a second  
recombination site and the heterologous nucleic acid;

(b) incubating the resulting mixture in the presence of at least one  
recombination protein under conditions whereby recombination between  
25 the first and second recombination sites is effected, thereby introducing  
the heterologous nucleic acid into the platform-ACes within the  
mesenchymal cell.

65. A lambda-intR mutein comprising a glutamic acid to arginine  
change at position 174 of wild-type lambda-intR.

66. The lambda-intR mutain of claim 65, wherein the lambda-intR mutain comprises SEQ ID NO:37.

67. The method of claim 46 wherein the promoterless marker is transcriptionally downstream of the heterologous nucleic acid, wherein  
5 the heterologous nucleic acid encodes a heterologous protein, and wherein the expression level of the selectable marker is transcriptionally linked to the expression level of the heterologous protein.

68. The method of claim 67, wherein the selectable marker and the heterologous nucleic acid are transcriptionally linked by the presence  
10 of a IRES between them.

69. The method of claim 68, wherein the selectable marker is selected from the group consisting of an antibiotic resistance gene, and a detectable protein, wherein the detectable protein is chromogenic or  
fluorescent.

70. The method of claim 69, wherein the selectable marker is selected from the group consisting of green fluorescent protein (GFP), red  
15 fluorescent protein (RFP), blue fluorescent protein (BFP), and *E. coli* histidinol dehydrogenase.

71. The method of claim 67 further comprising expressing the  
20 heterologous protein and isolating the heterologous protein.

72. A method for producing a transgenic animal, comprising introducing a platform-ACes into an embryonic cell.

73. The method of claim 72, wherein the embryonic cell is a stem cell.

74. The method of claim 72, wherein the embryonic cell is in an  
25 embryo.

75. The method of claim 72, wherein the platform-ACes comprises heterologous nucleic acid that encodes a therapeutic product.

76. The method of claim 72, wherein the transgenic animal is a fish, insect, reptile, amphibians, arachnid or mammal.

77. The method of claim 72, wherein the *ACes* is introduced by cell fusion, lipid-mediated transfection by a carrier system, microinjection,  
5 microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer.

78. A transgenic animal produced by the method of claim 72.

79. A cell line useful for making a library of *ACes*, comprising a multiplicity of heterologous recombination sites randomly integrated  
10 throughout the endogenous chromosomes.

80. A method of making a library of *ACes* comprising random portions of a genome, comprising introducing one or more *ACes* into the cell line of claim 79, under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of  
15 the heterologous recombination sites within the cell's chromosomal DNA; and isolating said multiplicity of *ACes*, thereby producing a library of *ACes* whereby multiple *ACes* have different portions of the genome within.

81. A library of cells useful for genomic screening, said library  
20 comprising a multiplicity of cells, wherein each cell comprises an *ACes* having a mutually exclusive portion of a chromosomal nucleic acid therein.

82. The library of cells of claim 81, wherein the cells of the library are from a different species than the chromosomal nucleic acid  
25 within the *ACes*.

83. A method of making one or more cell lines, comprising  
a) integrating into endogenous chromosomal DNA of a selected cell species, a multiplicity of heterologous recombination sites,

b) introducing a multiplicity of *ACes* under conditions that promote the site-specific chromosomal arm exchange of the *ACes* into, and out of, a multiplicity of the heterologous recombination sites integrated within the cell's endogenous chromosomal DNA;

5        c) isolating said multiplicity of *ACes*, thereby producing a library of *ACes* whereby a multiplicity of *ACes* have mutually exclusive portions of the endogenous chromosomal DNA therein;

d) introducing the isolated multiplicity of *ACes* of step c) into a multiplicity of cells, thereby creating a library of cells;

10       e) selecting different cells having mutually exclusive *ACes* therein and clonally expanding or differentiating said different cells into clonal cell cultures, thereby creating one or more cell lines.

84. The method of claim 23, wherein the nucleic acid molecule with a recombination site is a PCR product.

15       85. Method of claim 23 wherein the recombinase is a protein and the recombination event occurs in vitro.

86. The method of claim 33, wherein the vector is a PCR product comprising a second recombination site.

20       87. The lambda-intR mutein of claim 65, wherein the mutein further comprises an amino acid signal for nuclear localization.

88. The lambda-intR mutein of claim 65, wherein the mutein further comprises an epitope tag for protein purification.

89. A modified iron-induced promoter comprising SEQ ID NO:128.

25       90. A plasmid or expression cassette comprising the promoter of claim 89.

91. A vector, comprising:  
a recognition site for recombination; and

a sequence of nucleotides that targets the vector to an amplifiable region of a chromosome.

92. The vector of claim 91, wherein the amplifiable region comprises heterochromatic nucleic acid.

5 93. The vector of claim 91, wherein the amplifiable region comprises rDNA.

94. The vector of claim 93, wherein the rDNA comprises an intergenic spacer.

10 95. The vector of claim 91, further comprising nucleic acid encoding a selectable marker that is not operably associated with any promoter.

96. The vector of claim 91, wherein the chromosome is a mammalian chromosome.

15 97. The vector of claim 91, wherein the chromosome is a plant chromosome.

98. A cell of claim 57 that is a plant cell, wherein the ACes platform is a MAC.

99. The plant cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

20 100. The plant cell of claim 99, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

101. A cell of claim 57 that is an animal cell, wherein the ACes platform is a plant artificial chromosome (PAC).

25 102. The cell of claim 101 that is a mammalian cell.

103. The cell of claim 98, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

104. The cell of claim 102, wherein the MAC comprises transcriptional regulatory sequence of nucleotides derived from plants.

105. The cell of claim 104, wherein the regulatory sequence is selected from the group consisting of promoters, terminators, enhancers, silencers and transcription factor binding sites.

5 106. A method, comprising:  
introducing a vector of claim 91 into a cell;  
growing the cells; and  
selecting a cell comprising an artificial chromosome that comprises one or more repeat regions.

10 107. The method of claim 106, wherein sufficient portion of the vector integrates into a chromosome in the cell to result in amplification of chromosomal DNA.

108. The method of claim 106, wherein the artificial chromosome is an *ACes*.

15 109. A method for screening, comprising:  
contacting a cell comprising a reporter *ACes* with test compounds or known compounds, wherein:

the reporter *ACes* comprises one or a plurality of reporter constructs;

20 a reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds; and  
detecting any increase or decrease in signal output from the reporter, wherein a change in the signal is indicative of activity of the test or known compound on the regulatory region.

25 110. The method of claim 109, wherein the reporter is operatively linked to a promoter that controls expression of a gene in a signal transduction pathway, whereby activation or reduction in the signal indicates that the pathway is activated or down-regulated by the test compound.

111. The method of claim 109, wherein the reporter in the construct encodes drug resistance or encodes a fluorescent protein.

112. The method of claim 111, wherein the fluorescent protein is selected from the group consisting of red, green and blue fluorescent  
5 proteins.

113. The method of claim 109, wherein the *ACes* comprises a plurality of reporter-linked constructs, each with a different reporter, whereby the pathway(s) affected by the test compounds can be elucidated.

10 114. The method of claim 109, wherein a reporter is operatively linked to a promoter that is transcriptionally regulated in response to DNA damage, and the test compounds are genotoxicants.

115. The method of claim 114, wherein the DNA damage is induced by apoptosis, necrosis or cell-cycle perturbations.

15 116. The method of claim 114, wherein unknown compounds are screened to assess whether they are genotoxicants.

117. The method of claim 114, wherein the promoter is a cytochrome P450-profiled promoter.

20 118. The method of claim 114, wherein the cell is in a transgenic animal and toxicity is assessed in the animal.

119. The method of claim 109, wherein:

the cell is a patient cell sample; the patient has a disease;

the regulatory region is one targeted by a drug or drug regimen;

and

25 the method assesses the effectiveness of a treatment for the disease for the particular patient.

120. The method of claim 119, wherein the cell is a tumor cell.

121. The method of claim 109, wherein the cell is a stem cell or a progenitor cell, whereby expression of the reporter is operatively linked to

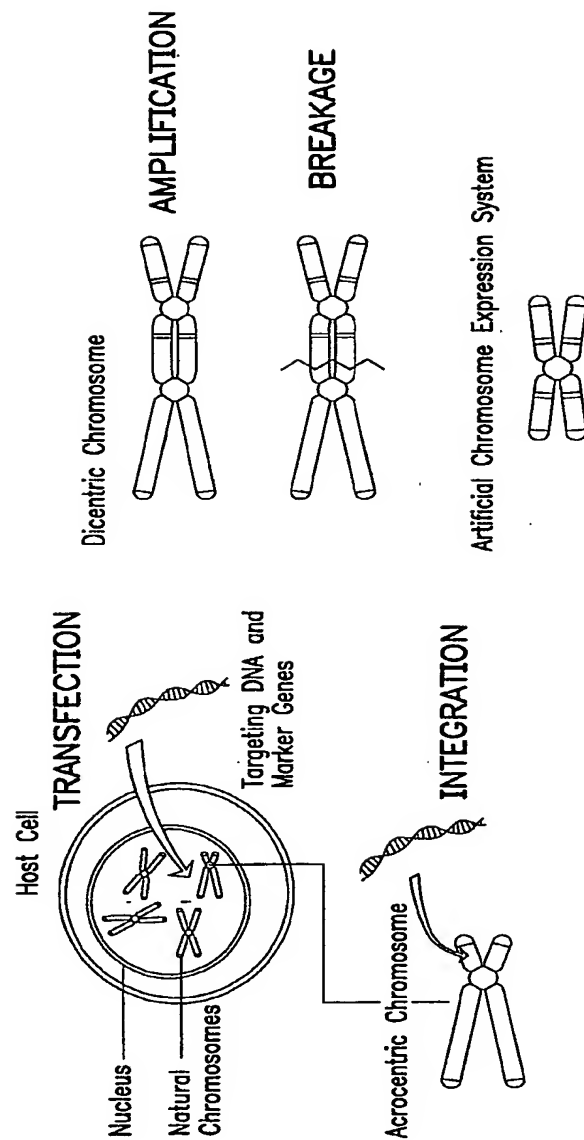
a regulatory region expressed in the cells to thereby identify stem cells or progenitor cell.

122. The method of claim 109, wherein the cell is in an animal;  
and the method comprises whole-body imaging to monitor expression of  
5 the reporter in the animal.

123. A reporter *ACes* comprises one or a plurality of reporter constructs, wherein the reporter construct comprises a reporter gene in operative linkage with a regulatory region responsive to test or known compounds.



# Generation of ACes for Platform Chromosome Engineering



**FIG. 1**

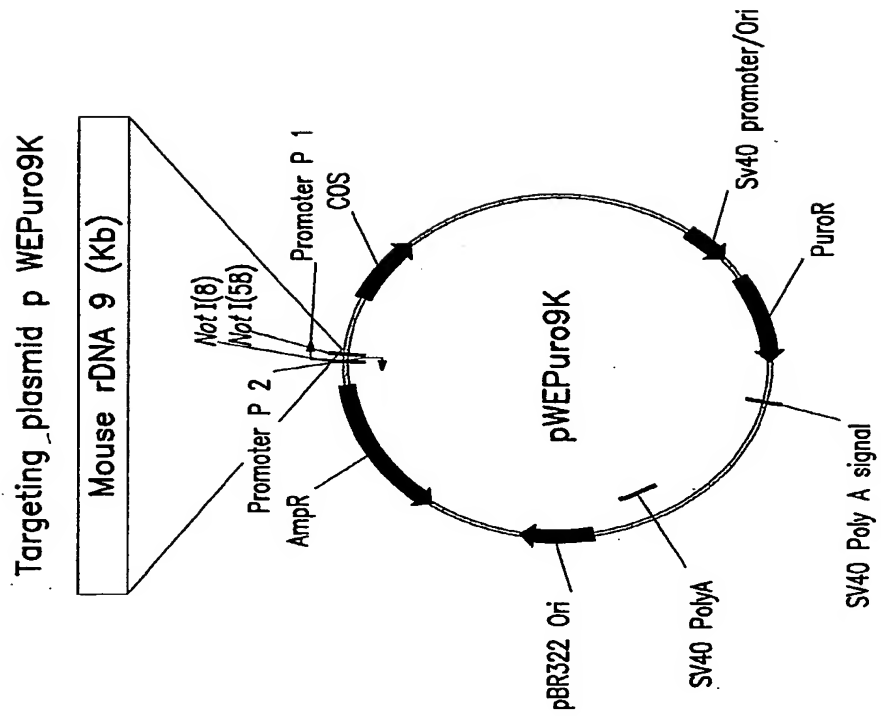


FIG. 2

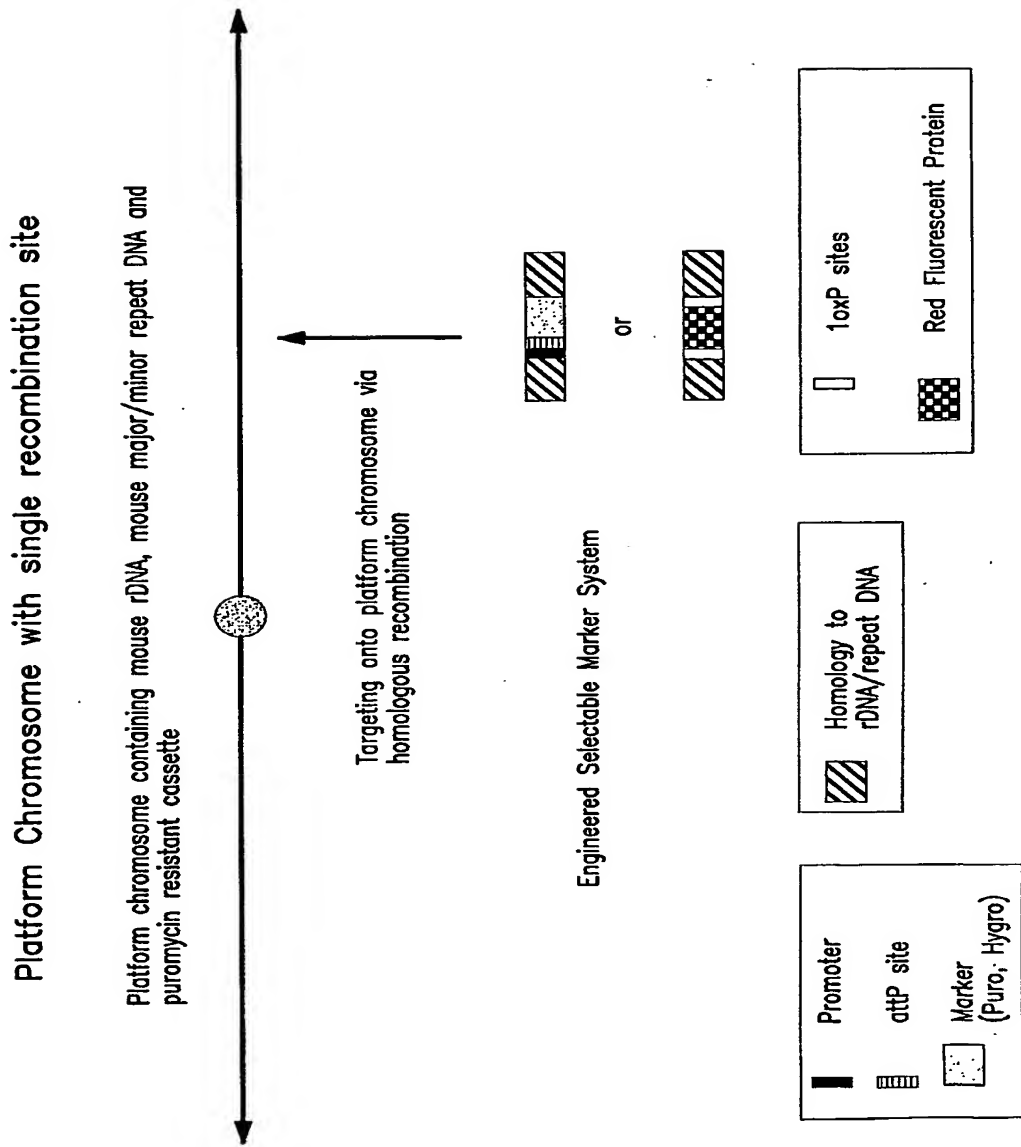


FIG. 3

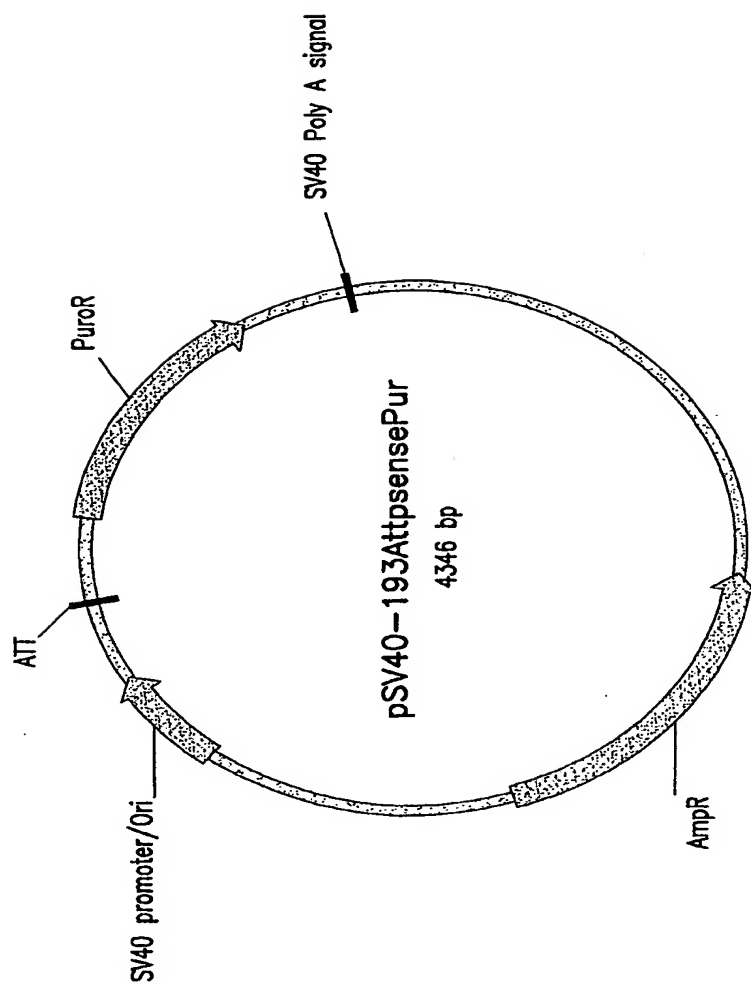


FIG. 4

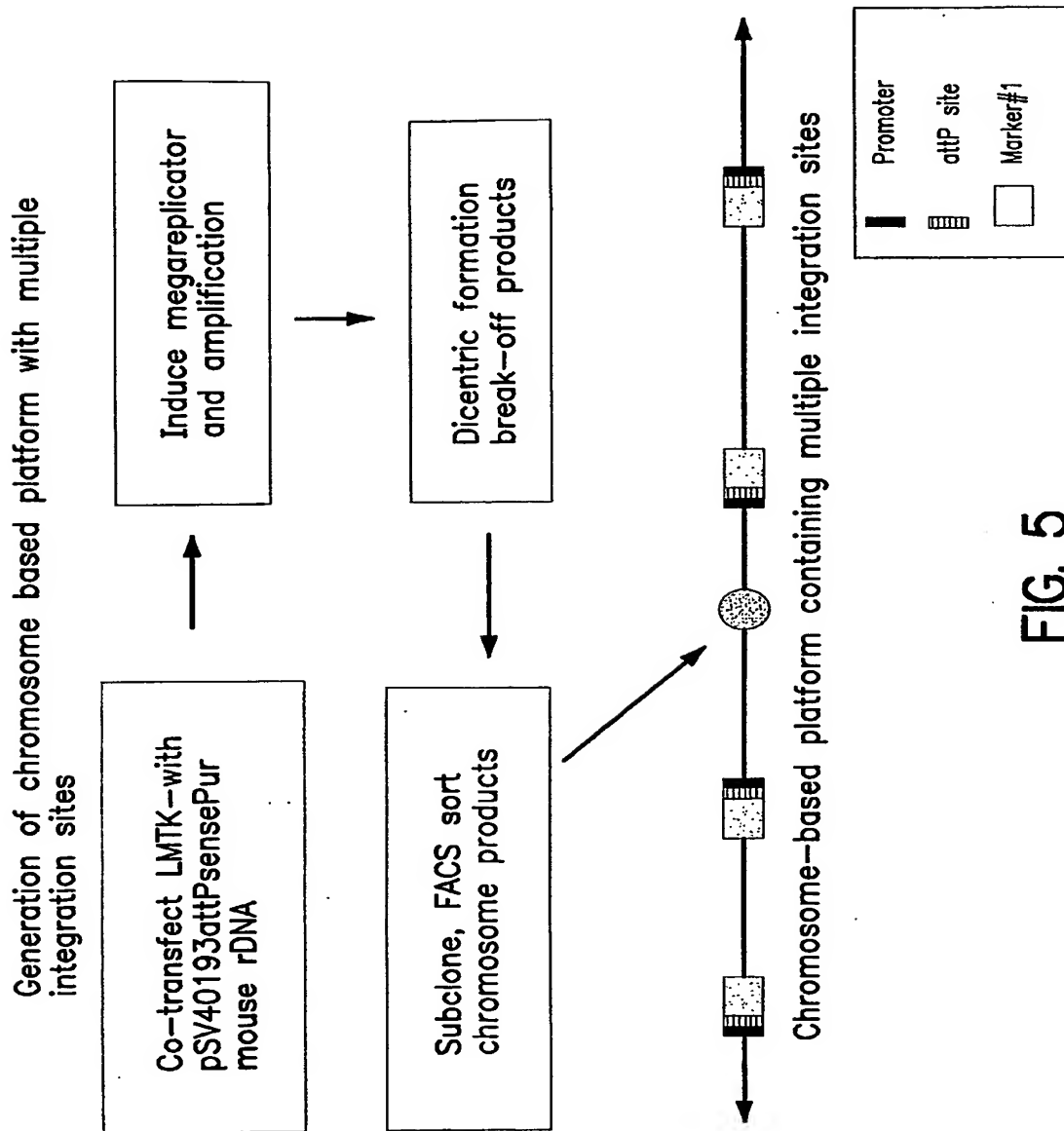


FIG. 5

$\lambda$  integrase recombination



	Core Region
attP	CAGCTTTT <u>T</u> TATACTAAGTTG
attB	CTGCTTTT <u>T</u> TATACTAACTTG
attL	CTGCTTTT <u>T</u> TATACTAAGTTG
attR	CAGCTTTT <u>T</u> TATACTAACTTG

FIG. 6

# $\lambda$ INT recombination on artificial chromosome

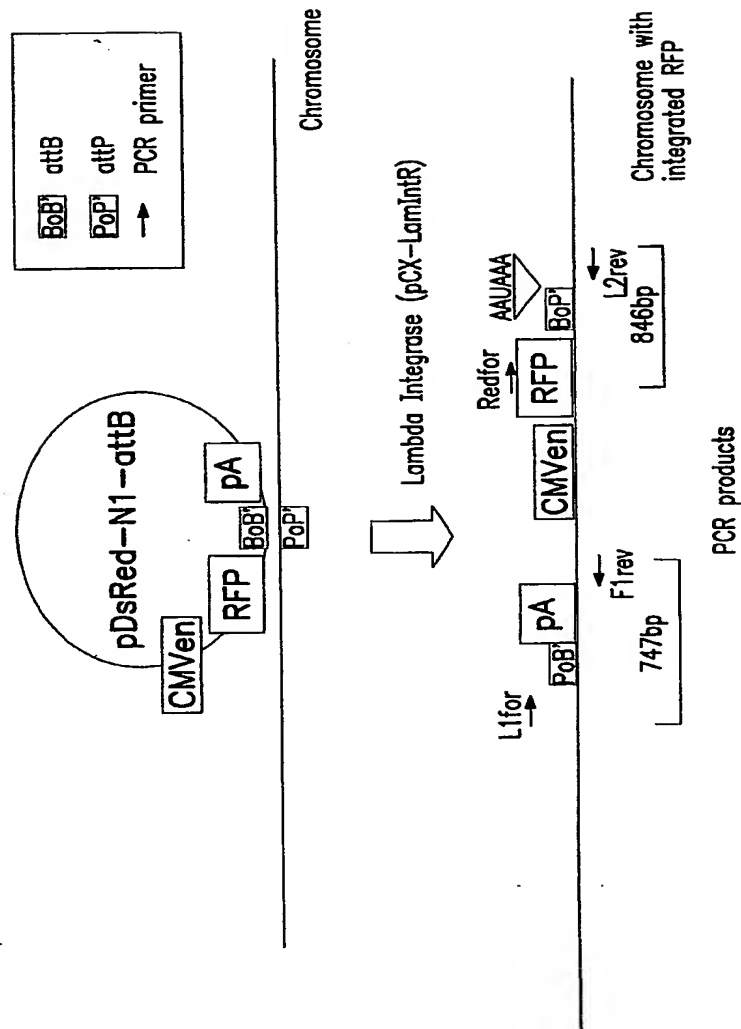


FIG. 7

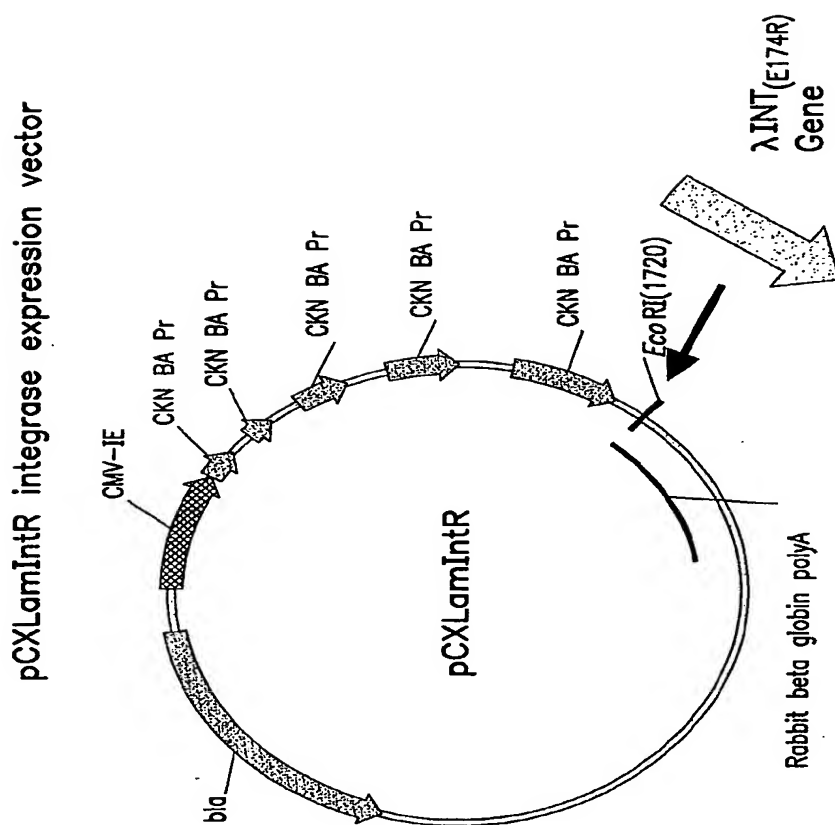


FIG. 8



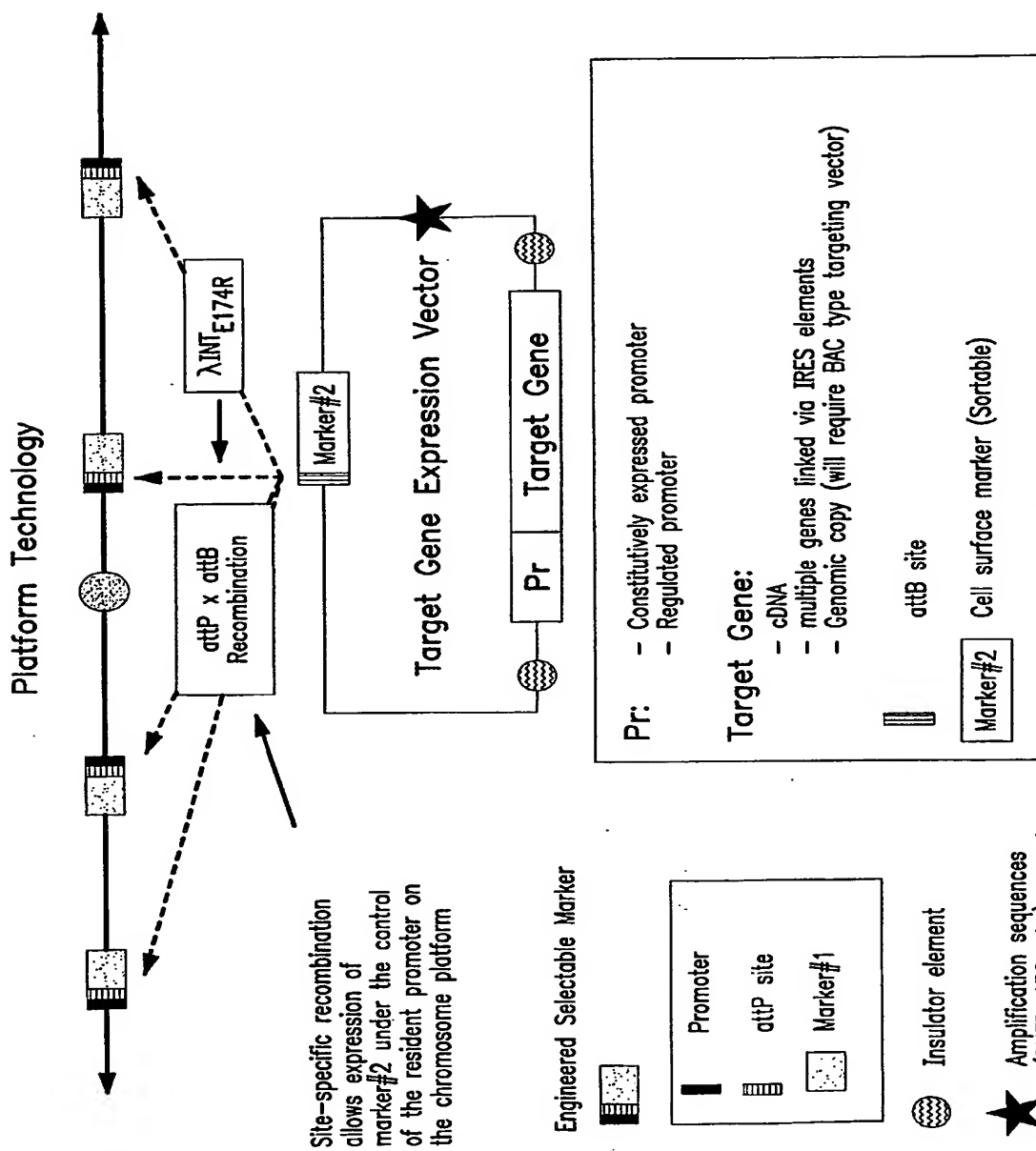


FIG. 9

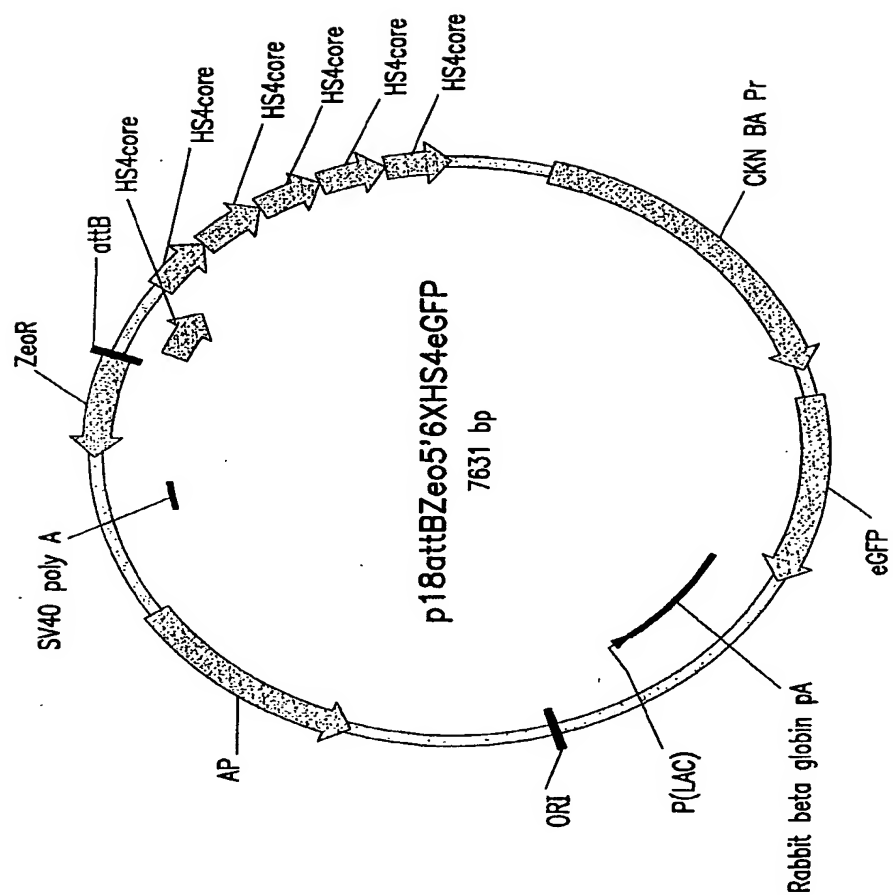


FIG. 10

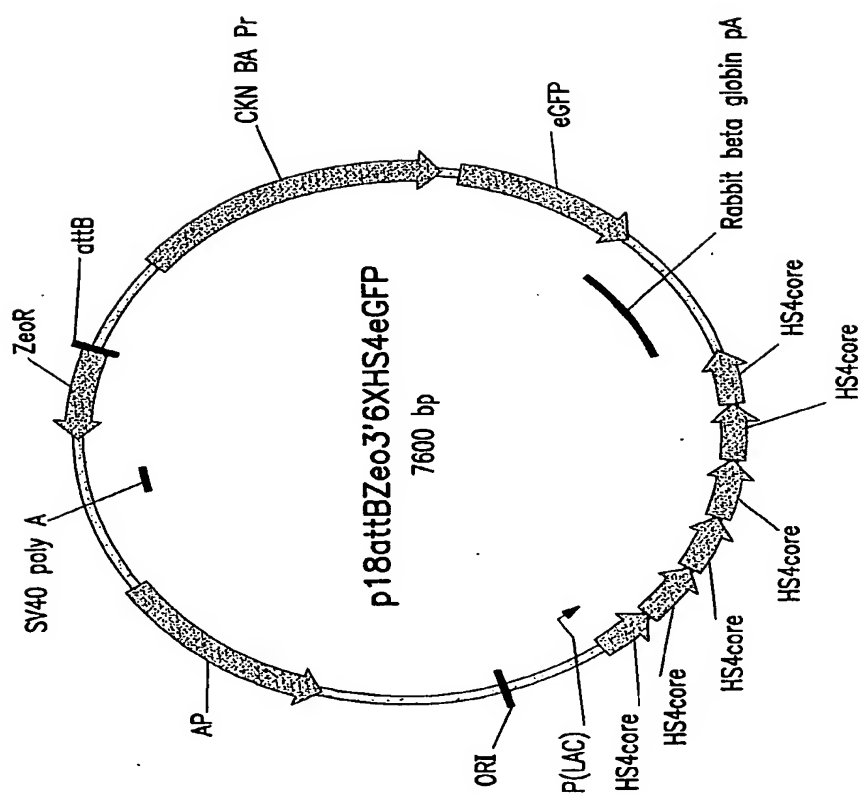


FIG. 11

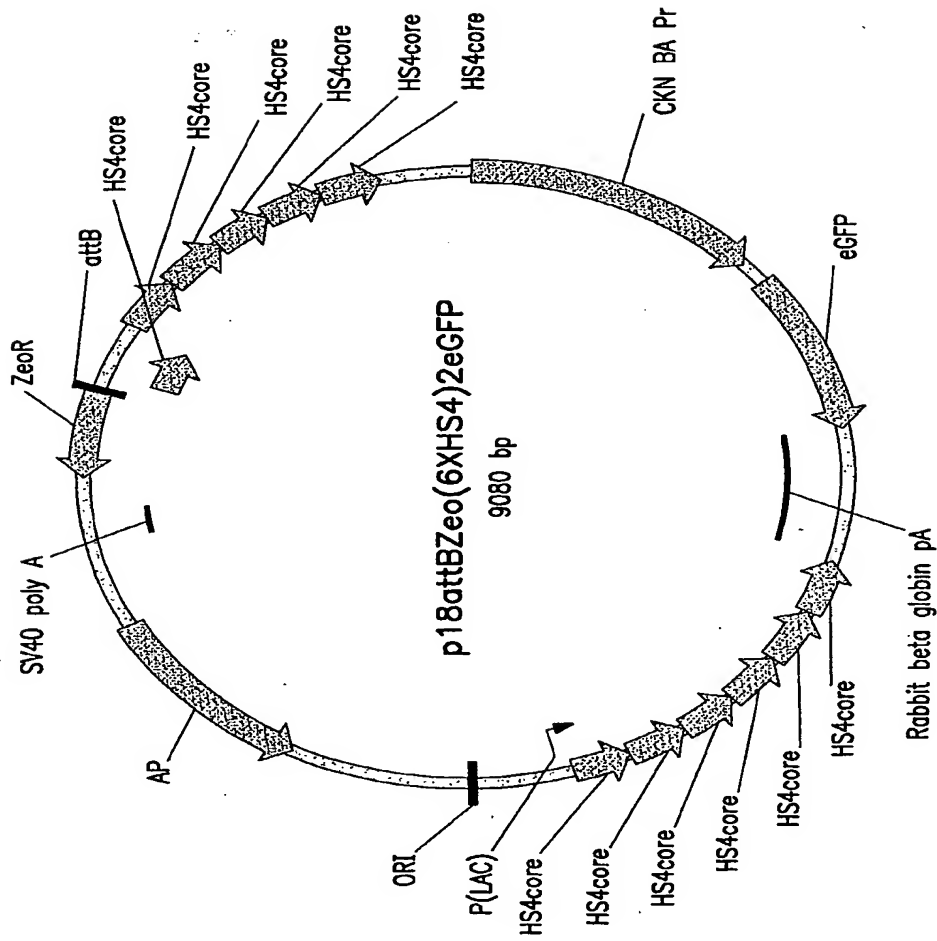


FIG. 12

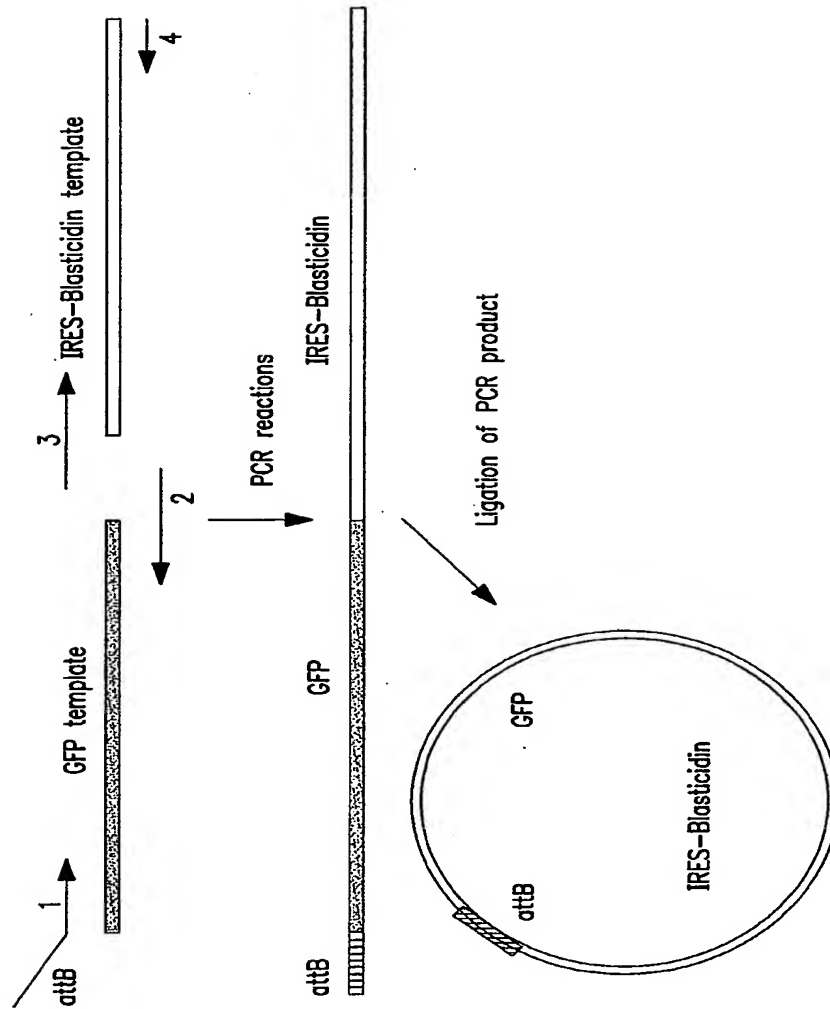


FIG. 13

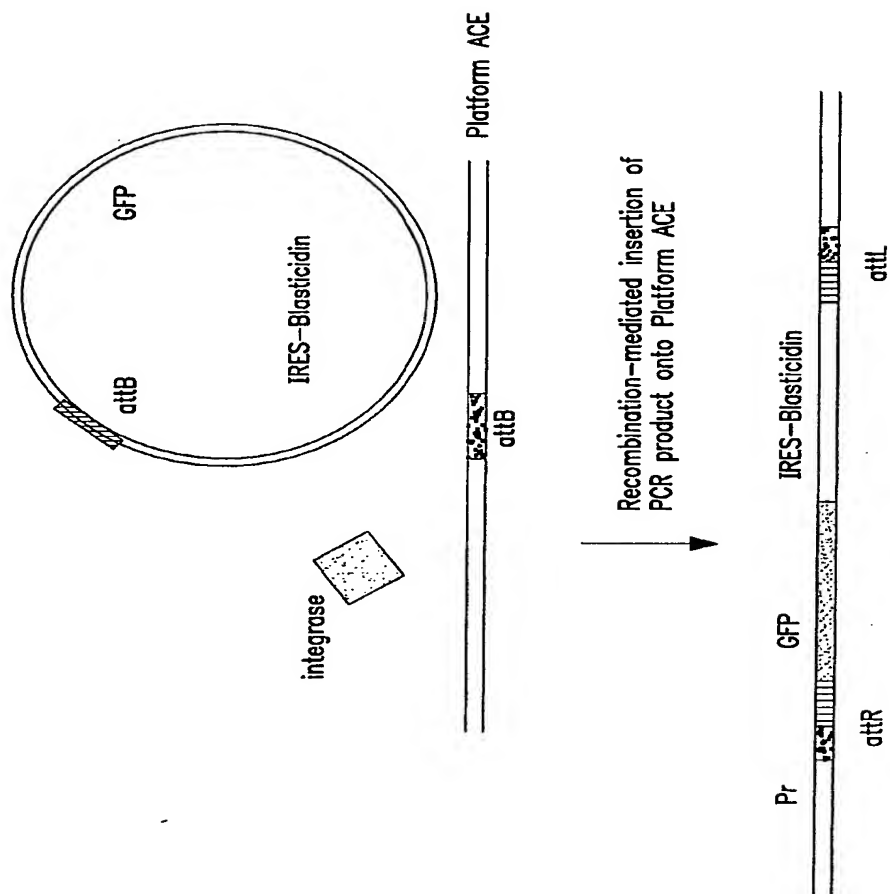


FIG. 14

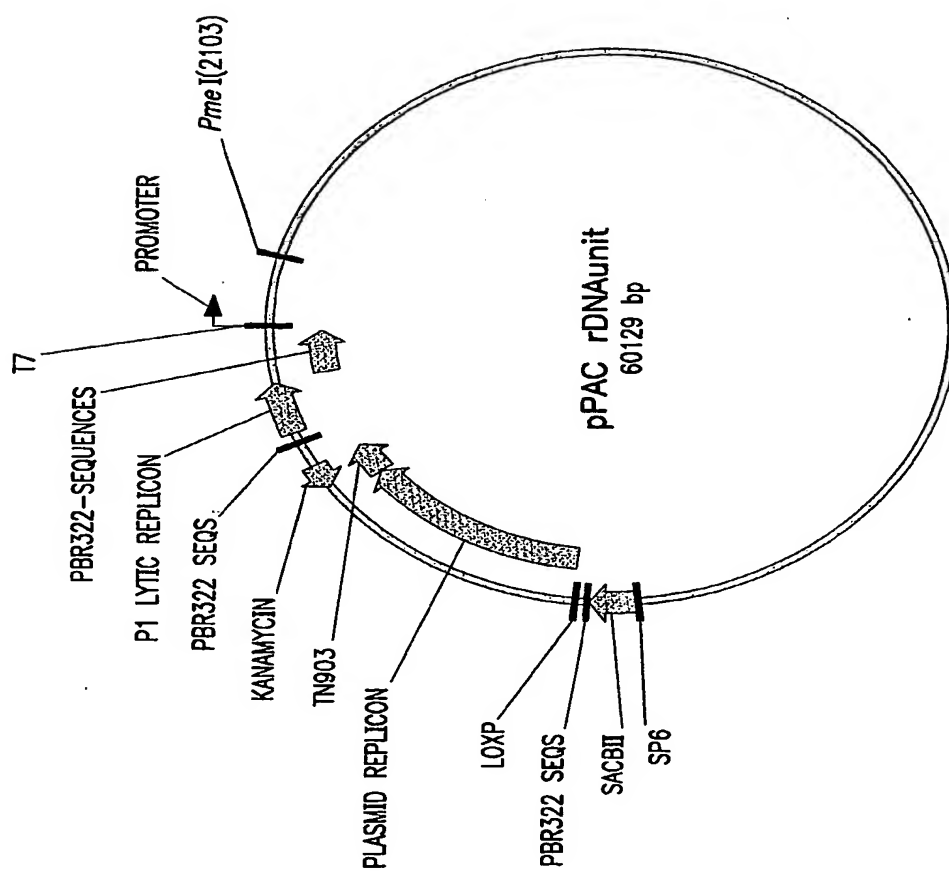


FIG. 15

SEQUENCE LISTING

<110> CHROMOS MOLECULAR SYSTEMS, INC.

Perkins, Edward  
Perez, Carl  
Lindenbaum, Michael  
Greene, Amy  
Leung, Josephine  
Fleming, Elena  
Stewart, Sandra  
Shellard, Joan

<120> CHROMOSOME-BASED PLATFORMS

<130> 24601-420PC

<140> Not Yet Assigned

<141> Herewith

<150> 60/294,758

<151> 2001-05-30

<150> 60/366,891

<151> 2002-03-21

<160> 129

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer: attPUP

<400> 1

ccttgcgcta atgctctgtt acagg

25

<210> 2

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer: attPDWN

<400> 2

cagaggcagg gagtgggaca aaattg

26

<210> 3

<211> 35

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer: Lamint 1

<400> 3

ttcgaattca tgggaagaag gcgaagtcac gagcg

35

<210> 4

<211> 34

<212> DNA

<213> Artificial Sequence



<220>  
 <223> Primer: Lamint 2

<400> 4  
 ttctgaattct tatttgattt caattttgtc ccac 34

<210> 5  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer

<400> 5  
 cggacaatgc ggttgtgcgt 20

<210> 6  
 <211> 46  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> primer

<400> 6  
 cgcgcagcaa aatctagagt aaggagatca agacttacgg ctgacg 46

<210> 7  
 <211> 46  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> LambdaINTER174rev

<400> 7  
 cgtcagccgt aagtcttgat ctcttactc tagattttgc tgcgcg 46

<210> 8  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> attB1

<400> 8  
 tgaagcctgc ttttttatac taacttgagc gaa 33

<210> 9  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> attB2

<400> 9  
 ttctgtcaag ttagtataaa aaagcaggct tca 33

<210> 10  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

<220>

<223> Primer: attPdwn2

<400> 10  
tcttctcggg cataagtcgg acacc 25

<210> 11  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer:CMVen

<400> 11  
ctcacgggga tttccaagtc tccac 25

<210> 12  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer:attPdwn

<400> 12  
cagaggcagg gagtgggaca aaattg 26

<210> 13  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer:CMVEN2

<400> 13  
caactccgcc ccattgacgc aaatg 25

<210> 14  
<211> 26  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer:L1

<400> 14  
agtatcgccg aacgattagc tcttca 26

<210> 15  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer:F1 rev

<400> 15  
gccgatttcg gcctattggt taaa 24

<210> 16  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer:RED

<400> 16  
ccgccgacat ccccgactac aagaa

25

<210> 17  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer:L2rev

<400> 17  
ttccttcgaa ggggatccgc ctacc

25

<210> 18  
<211> 22118  
<212> DNA  
<213> Mus musculus

<300>  
<308> GenBank X82564  
<309> 1996-04-09

<400> 18	gaattcccct	atcccctaate	cagattgggtg	gaataaacttg	gtatagatgt	ttgtgcatta	60
	aaaaccctgt	aggatcttca	ctctaggtca	ctgttcagca	ctggaacctg	aattgtggcc	120
	ctgagtgata	ggctcctggga	catatgcagt	tctgcacaga	cagacagaca	gacagacaga	180
	cagacagaca	gacagacgtt	acaaacaaac	acgttgagcc	gtgtgccaac	acacacacaa	240
	acaccactct	ggccataaatt	attgaggacg	ttgatttatt	attctgtgtt	tgtgagtcgt	300
	tctgtctgtc	tgtctgtctg	tctgtctgtc	tatcaaacca	aaagaaacca	aacaattatg	360
	cctgcctgcc	tgcctgcctg	cctacacaga	gaaatgattt	cttcaatcaa	tctaaaacga	420
	cctcctaagt	ttgccttttt	tctctttctt	tatcttttct	ttttttcttt	tcttcttctt	480
	tcttctcttc	cttcttctct	tcttctcttt	ctttctttct	ttctttcttt	cttactttct	540
	ttctttctct	cttacatttta	ttcttttctt	acatagtttc	ttagtgttaag	catccctgac	600
	tgtcttgaag	acacttttga	ggcctcaatc	ctgtaagagc	cttctctctg	ttttcaaatg	660
	ctggcatgaa	tgttgtagct	cactatgacc	agcttagtct	tcaagtctga	gttactggaa	720
	aggagttcca	agaagactgg	ttatattttt	catttattat	tgcattttta	ttaaaattta	780
	atttcaccaa	aataggtcac	actgaccaat	ctcagatctg	ccgtttaaaa	gcataaggaa	840
	aaagtaggag	aaaacacgtg	ggctgtctgt	ggatggctga	ggctgcttta	gggagcctcg	900
	tcaccattct	gcacttgcaa	accggggcac	tagaaccggg	tgaaggagga	aaccaaagcg	960
	acctggaaac	aataggtcac	atgaaggcca	gccacctcca	tcttggtgtg	cgggagttca	1020
	gttagcagac	aagatggctg	ccatgcacat	gttggtcttc	agcttggtga	gggtcaaagta	1080
	caaccgagtc	acagaacaag	gaagtataca	cagtgagttc	caggtcagcc	agagtttaca	1140
	cagagaaaac	acatcttgaa	aaaaacaaaa	aaataaatta	aataaatata	atttaaaaat	1200
	ttaaaaatag	ccgggagtg	tggcgcatgt	ctttaatccc	agctctcttc	aggcagagat	1260
	gggaggattt	ctgagtttga	ggccagcctg	gtctgcaaag	tgagttccag	gacagtccag	1320
	gctatacaga	gaaaccctgt	cttgaaaact	aaactaaatt	aaactaaaat	aaactaaaaa	1380
	aatataaaat	aaaaatttta	aagaatttta	aaaaactaca	gaaatcaaac	ataagcccac	1440
	gagatggcaa	gtaactgcaa	tcatagcaga	aatattatac	acacacacac	acacagactc	1500
	tgtcataaaa	tccaatgtgc	cttcatgatg	atcaaatttc	gatagtcagt	aatactagaa	1560
	gaatcatatg	tctgaaaata	aaagccagaa	ccttttctgc	ttttgttttc	ttttgcccac	1620
	agatagggtt	tctctcagtg	tatccctggc	atccctgcct	ggaaacttct	ttgtagggtt	1680
	ggtagcctca	aactcagaga	ggctcctctc	gcctgcctgc	ctgcctgcct	gcctgcctgc	1740
	ctgcctgcct	gcctgcctca	cttcttctgc	cacccacaca	accgagtcga	acctaggatc	1800
	tttattttct	tctctttctc	tcttctttct	ttctttcttt	ctttctttct	ttctttcttt	1860
	ctttctttct	cttctatttc	attagttttc	aatgtaagtg	tgtgtttgtg	ctctatctgc	1920
	tgcctatagg	cctgcttgcc	aggagagggc	aacagaacct	aggagaaacc	accatgcagc	1980
	tcctgagaat	aagtgaaaaa	acaacaaaaa	aaggaaattc	taatcacata	gaatgtagat	2040
	atatgccgag	gctgtcagag	tgccttttaa	ggcttagtgt	aagtaaatga	aattgttgtg	2100
	tgtcttttat	ccaaacacag	aagagaggtg	gctcggcctg	catgtctgtt	gtctgcattg	2160
	agaccaggct	ggccttgaac	acattaatct	gtctgcctct	gcttccctaa	tgtctgcatt	2220
	aaaggcctgt	gccaccactg	cccggactga	tttcttcttt	tttttttttt	tggaaaatac	2280
	ctttctttct	tttcttctct	cttcttctct	ttcttctctt	ttcttctctt	tttttttttt	2340
	tttctttttt	cttttttttt	ttttttttta	aatttgccta	aggttaaagg	tgtgtctccac	2400
	aattgcctca	gctctgctct	aattctcttt	aaaaaaaaaa	aaacaaaaaa	aaaacaaaaa	2460
	cagtatgtat	gtatgtatat	ttagaagaaa	tactaatcca	ttaataaact	ttttttctta	2520
	aaattcatgt	cattcttgtt	ccacaaagtg	agttccagga	cttaccagag	aaacctgtgt	2580

ttcaaat	ttc	tgtgttcaag	gtcaccctgg	cttacaaagt	gagttccaag	tccgataggg	2640
ctacacagaa	aaaccat	atc	tcagaaaaa	aaaaagttcc	aaacacacac	acacacacac	2700
acacacacac	acacacacac	acacacacac	acacacacac	acacacacag	cgcgcccgcg	cgatgagggg	2760
aagtcgtgcc	taaaataaat	at	tttttctgg	ccaaagtga	agcaaatac	tatgaagagg	2820
tactcctaga	aaaaataaat	acaaacgggc	tttttaataca	ttccagcact	gttttaattt	gttttaattt	2880
aactctgaat	ttagtcttgg	aaaagggggc	gggtgtgggt	gagtgagggc	gagcgagcag	gagcgagcag	2940
acggggcgggc	ggggcggtga	gtggccggcg	gcggtggcag	cgagcaccag	aaaacaacaa	aaaacaacaa	3000
accccaagcg	gtagagtgtt	ttaaaaatga	gacctaaatg	tggtggaacg	gaggtcgccg	gaggtcgccg	3060
ccaccctcct	cttccactgc	ttagatgtct	ccttcccctt	actgtgctcc	cttcccctaa	cttcccctaa	3120
ctgtgcctaa	ctgtgcctgt	tccctcacc	cgctgattcg	ccagcgacgt	actttgactt	actttgactt	3180
caagaacgat	tttgccctgtt	ttcaccgct	cctgtcatac	tttcgttttt	gggtgcccga	gggtgcccga	3240
gtctagcccg	ttcgctatgt	tccggcgggg	cgatggggac	cgtttgtgtcc	actcgggaga	actcgggaga	3300
agtgggtggg	gggtacgctg	ctccgctcgt	cggtcgtgag	tgagctcggg	gatctgtatt	gatctgtatt	3360
agacctccg	gagagacaga	atgagtga	gaatgtggcg	gcgcgtgacg	agtttcggga	agtttcggga	3420
ggtttgtatg	gttgatcgag	accattgtcg	ggcgacacct	agtggtgaca	tgagcgac	tgagcgac	3480
acgctccagg	cctctcaggt	tggtgacaca	ggagaggga	gtgcctgtgg	cgtgtttaag	cgtgtttaag	3540
agggtgacag	gaggccggcg	aagcaggcgg	gagcgtctcg	gagatgggtg	gttgctgtac	gttgctgtac	3600
gacgggtctct	aacaaggagg	tcgtacaggg	agatggccaa	agcagaccga	agaaggctta	agaaggctta	3660
gcccttttgg	gaaaaatgct	aggggtgggt	gcaacgttac	taggtcgacc	ttgtcccctg	ttgtcccctg	3720
agtcctaccc	ccccccccct	tttttttttt	tttccctcag	aagccctctc	tccaagcccg	tccaagcccg	3780
caccggggggc	accgtacatc	tgaggccgag	aggacgcgat	gggcccggct	tttccctcca	tttccctcca	3840
tgtggctcgg	ccagctggcg	cttcgggtct	tttttttttt	tttttttttt	aggacgcgat	aggacgcgat	3900
gaagccttgt	ctgtcgctgt	caccggggcg	gctgtacttc	tgaggccgag	tttttttttt	tttttttttt	3960
gggccccggg	ttccaagccg	gtgtggctcg	gccagctgga	gcttcgggtc	gtcaccgggg	gtcaccgggg	4020
tttttttttt	ttttttttct	cagaagcctt	gtctgtcgct	gtcaccgggg	gccagctgga	gccagctgga	4080
tctgaggccg	agaggacgcg	atgggtcggc	ttccaagccg	atgtggcggg	gggggagcgt	gggggagcgt	4140
gcttcgggtt	tttttttttc	ctccagaagc	cctctcttgt	ccccgtcacc	gcccgggtcag	gcccgggtcag	4200
tacttctgag	gccgagagga	cgtgatgggc	cggggttcca	ggcggatgtc	tcccgcctac	tcccgcctac	4260
ctggagcctt	ggatcttttt	tttttttttt	cctccagaag	ccctctcttg	aagccgatgt	aagccgatgt	4320
cggggggcacc	ttacatctga	gggcgagagg	acgtgatggg	tccggcttcc	ctcttgtccc	ctcttgtccc	4380
ggcggggcca	gctggagctt	cgggtttttt	ttttttcttc	cagaagccct	gggtcccagg	gggtcccagg	4440
cgtcaccggg	ggcgtgtac	ttctgagcc	gagaggacct	gatgggcccg	agaagccctc	agaagccctc	4500
ggatgtcgcc	cggtcagctg	gcaccgtaca	tcattttttt	ttttccctcc	atggggcctgt	atggggcctgt	4560
tcttgtcccc	gtcaccgggg	ggccagctgg	tcctgaggcc	agaggacacg	ttttttcttc	ttttttcttc	4620
cttccaagcc	gatgtggccc	ggccagctgg	agcttcgggt	cctttttttt	agaggacgcg	agaggacgcg	4680
cagaagcctt	gtctgtcgct	gtcaccgggg	gcgctgtact	tctgaggccg	cctttttttt	cctttttttt	4740
atggggcccg	cttccaagcc	gggtgtggctc	ggccagctgg	agcttcgggt	ttgtacttct	ttgtacttct	4800
tttttttttt	ttccctccaga	aaccttgtct	tcctgagctg	cccggggcgc	cagctggagc	cagctggagc	4860
gatgccgaga	ggacgcgatg	ggcccgctct	ccaggccgat	gtggcccggt	tcaccggggg	tcaccggggg	4920
tttggatctt	tttttttttt	ttttccctcca	gaagccctct	cctgtccccg	atgtggcccg	atgtggcccg	4980
caccttacat	ctgaggcccta	gaggacacga	tgggcccggg	ttccaggccg	tgtccccctc	tgtccccctc	5040
gtcagctgga	gctttggatc	tttttttttt	ttttcttcca	gaagccctct	ccaatccgat	ccaatccgat	5100
accggtggca	ctgtacatct	gaggcgagga	ggacattatg	ggcccggcct	ttccagaagc	ttccagaagc	5160
gtggcccggg	cagctggagc	tttggatctt	atattttttt	taattttttt	ttatgggccc	ttatgggccc	5220
cctcttgtcc	ctgtcacccg	ttggcaggta	catctgaggg	cgagaggaca	ttttttttct	ttttttttct	5280
ggcttccagg	ccgatgtggc	ccggtcagct	ggagctttgg	atcttttttt	ttttttttct	ttttttttct	5340
tttttccctc	agaagccctc	tctgtcccctg	tcaccggggg	ccctgtacgt	ctgaggccga	ctgaggccga	5400
gggaaagcta	tgggcgcggg	tttctttcat	tgacctgtcg	gtcttatcag	ttctccgggt	ttctccgggt	5460
tgtcagggtc	gaccagttgt	tccttttgagg	tcceggttctt	ttcgttatgg	gggtcatttt	gggtcatttt	5520
gggccacctc	cccaggtatg	acttccaggc	gtcgttgctc	gtcgttgctc	ttcctccctg	ttcctccctg	5580
tctcttttat	gcttgtgato	ttttctatct	gttccctattg	gacctggaga	taggtactga	taggtactga	5640
cacgctgtcc	tttccctatt	aacactaaag	gacactataa	agagaccctt	tcgatttaag	tcgatttaag	5700
gctgttttgc	ttgtccagcc	tattcttttt	actggcttgg	gtctgtcgcg	gtgacctgaag	gtgacctgaag	5760
ctgtccccga	gccacgcttc	ctgctttccc	gggcttgctg	ccttgctgtg	cttgctgtgg	cttgctgtgg	5820
gcagcttgtg	acaactgggc	gctgtgactt	tgctgcgtgt	cagacgtttt	tcccgatctc	tcccgatctc	5880
cccagaggtg	cgttgtcaca	cctgtcccgg	ttggaatggt	ggagccagct	gtgggtgagg	gtgggtgagg	5940
gccaccttat	ttcggctcac	tttttttttt	ttttttcttc	ttggagtccc	gaacctccgc	gaacctccgc	6000
tcttttctct	tcccgggtctt	tcttccacat	gcctcccagc	tgcattttct	tttgtttttt	tttgtttttt	6060
ttcttttttt	tttttttttt	ttggggaggt	ggagagtccc	gagtagctca	ctcctgtctg	ctcctgtctg	6120
tggltgtccaa	gtgttccatg	cacgtgcctc	ccgagtgcac	ttttttttgt	ggcagctcgt	ggcagctcgt	6180
cgttgtgttc	tcttgtttctg	tgtctgcccg	tatcagtaac	tgtcttgccc	cgcgtgtaag	cgcgtgtaag	6240
acattctcat	ctcgttgtgt	tctcccagat	gcgcgtcgtt	gctcactctt	agatcgatgt	agatcgatgt	6300
gggtgcctcg	agttctcttc	gggcccaggc	caagccgcgc	caggcgaggg	acggacattc	acggacattc	6360
atggcgaaatg	gcgccgcgtc	ttctcgttct	gccagcgggc	cctcgtctct	ccaccccato	ccaccccato	6420
cgtctgcggg	tggtgtgtgg	aaggcagggg	tgcggctctc	cggcccagcg	ctgcccgcgg	ctgcccgcgg	6480
cgcacttttc	tcagtggttc	gcgtggtcct	tggtgatgtg	tgaggcgccc	ggttgtgccc	ggttgtgccc	6540
tcacgtgttt	cactttgggc	gtgtctcgtc	tgaccatgtt	cccagagtcg	gtggatgtgg	gtggatgtgg	6600

ccggtggcgt	tgcataccct	tcccgtctgg	tgtgtgcacg	cgctgtttct	tgtaagcgtc	6660
gaggtgctcc	tggagcgctc	caggtttgtc	tcctaggtgc	ctgcttctga	gctgggtggg	6720
gcgtcccca	ttccctgggtg	tgcctccggt	gctccgtctg	gctgtgtgccc	ttcccggttg	6780
tgtctgagaa	gcccgtgaga	gggggggtcga	ggagagaagg	aggggcaaga	cccccttct	6840
tcgtcgggtg	aggcgccac	cccgcgacta	gtacgcctgt	gcgtagggtc	ggtgctgagc	6900
ggtcgcgggt	gggggttgaa	agtttctcga	gagactcatt	gctttcccg	ggggagcttt	6960
gagaggcctg	gctttcgggg	gggaccggtt	gcagggtctc	ccctgtccgc	ggatgctcag	7020
aatgcccttg	gaagagaacc	ttcctgttgc	cgcagacccc	cccgcgcggt	cgcccgcggtg	7080
ttggtcttct	ggtttccctg	tgtgctcgte	gcatgcatcc	tctctcggtg	gccgggggtc	7140
gtcgggggtt	tgggtccgtc	ccgcccctcag	tgagaaaagt	tccttctcta	gctatcttcc	7200
ggaaagggtg	cgggcttctt	acggtctcga	ggggtctctc	ccgaatgggtc	ccctggagggtg	7260
ctcgccccct	gaccgcctcc	cgcgcgcgca	gcgtttggtc	tctcgtctac	cgcgggcccgc	7320
ggcctcccc	ctccgagttc	ggggagggtat	cacgcggggtc	agagcctgtc	tgctgctcgt	7380
ccgttgctgc	ggagcatgtg	gctcgggttg	tgtggttggg	ggctgggggag	agggctccgt	7440
gcacaccccc	gcgtgcgcgt	actttcctcc	cctcctgagg	gcccgcgtgc	ggacgggggtg	7500
tgggtaggcg	acgggtgggt	cccggttccc	caccgtctct	cccggtgctc	accggtgctc	7560
tcgctcgctg	gcgtccctct	cgctcgcgtc	cacgactttg	gcccgtcccg	cgacggcggtc	7620
ctgcccgcgc	cgtggtgctg	gctgtgtgct	tctcgggtctg	tgtggttgtg	tcgctcgcgc	7680
cccccttccc	cgcggcagcg	ttcccacggc	tggcgaaatc	gcgggaggtcc	tccttcccct	7740
cctcgggggtc	gagagggtcc	gtgtctggcg	ttgattgatc	tcgctctcgg	ggacggggacc	7800
gttctgtggg	agaacggctg	ttggccgcgt	ccggcgcgac	gtcggacgtg	gggacccact	7860
gcccgtcggg	ggtcttctgc	ggtaggcatc	ggtgtgtcgg	catcggtctc	tcctcgtgtg	7920
cggtgtcggc	tcctcgggct	cccggggggc	cgctcgtgtt	cggttcgggt	cggcgctgca	7980
ggtgtggtgg	gactgctcag	gggagtgggtg	cagtgtgatt	cccgccgggt	ttgctcgcgc	8040
ggccctgacc	ggtccgacgc	ccgagcgggtc	tctcgggtccc	ttgtgaggac	ccccctcccg	8100
gagggccccg	tttcggccgc	ccttgccgtc	gtcgcggggtc	ctcgttctgc	tgtgtcgttc	8160
ccccctcccc	gctcgcgcga	gcccgtcttt	tttccctctc	ccccccctct	cctctgactg	8220
accggtggcc	gtgctgtcgg	accccccgca	tggggggcggt	cgggcacgta	cgcgtccggg	8280
cggtcaccgg	gggggcccag	gggtaagaaa	gggtaagaaa	gtcgggtcgg	cgggcgggag	8340
gagctgtggg	ttggaggggc	tcocggcccc	gcccgcgtgg	cggtgtcttg	cgcgggtcttg	8400
gagagggctg	cgtgcgaggg	gaaaagggtg	ccccgcgagg	gcaaaggga	agaggctagc	8460
agtggctcatt	gtcccagcgg	tgtggtgggtc	tggtggccga	gggtgcgtctg	gggggctcgt	8520
ccggccctgt	cgtccgtcgg	gaaggccgct	ggtggggcct	gcccggagtgc	cgagggtgggt	8580
accctggcgg	tgggattaac	cccgccgcgc	tgtcccgggt	tggcggtggg	ggctccgggtc	8640
gatgtctacc	tcctcctccc	cgagggtcca	ggccttctcc	gcccggggtc	tcggccctcc	8700
cctcgttctc	ccctctcgcg	gggttcaagt	cgctcgtcga	cctccccctc	tcctccttcc	8760
catctctcgc	gcaatggcgc	cgcccaggtt	cacgggtgggt	tcgtcctccg	cctccgcttc	8820
tcgcccgggg	gtctctggcg	tcgggtctct	cctgcccga	ccccgttggc	gtgggtctct	8880
ctcgccgggt	tcgcccactc	ctggcttcgc	ccggagggtc	agggggcttc	ccgggtcccc	8940
gacgttgccg	ctcgcgtcgt	tgtgcttggg	ggggggccgc	tgccggcctc	gcccgcgcgt	9000
gagcccctgc	cgcacccgcc	gggtgtcggt	ttcgcgcgc	ggtcagtgtg	gcccctggcg	9060
tgtgtcgcgt	cgggagcgtg	tcgcctcgc	ggcggttaga	cgccgggtgtc	gcccgggtcc	9120
gacgggtggc	ctatccaggg	ctcgcctccg	cgacccccgc	cctgcccgtc	ccgggtgggtg	9180
tcgttgggtg	ggggagtga	tgggtctacc	gggtcattccc	tcocgcgtgg	tttgactgtc	9240
tcgcccgggt	cgcccttctc	tttccgccaa	ccccacgccc	aacccaccac	cctgctctcc	9300
cgcccgggtg	cggtcgacgt	tcgggtctct	ccgatgccga	gggggttcggg	atttgtgccc	9360
gggacggagg	ggagagcggg	taagagaggt	gtcggagagc	tgtcccgggg	cgacgctcgg	9420
ggtggctttg	ccgcgtcggt	gtgctcgcgg	acgggttttg	tcggaccccc	acgggggtcgg	9480
tcggcccgca	tgcactctcc	cgttccgcgc	gagcccccgc	ccgggtcacc	cccggtttgt	9540
cctcccgcga	ggctctccgc	cgcgcgcgc	tcctcctcct	ctctcgcgct	ctctgtccc	9600
cctgggtcctg	tcacaccccc	gacgtccgc	tcgcgcttcc	ttacctgggt	gatcctgcc	9660
ggtagcatat	gcttgtctca	aagattaagc	catgcatgtc	taagtacgca	cggccgggtac	9720
agtgaactg	cgaatggctc	attaaatcag	ttatggttcc	tttggctcgt	cgtcctctc	9780
ctacttggat	aactgtggta	attctagagc	taatacatgc	cgacgggggc	tgacccccct	9840
tcocgggggg	ggatgctgct	atttatcaga	tcaaaaccaa	cccggtgagc	tcctccccgg	9900
ctccggccgg	gggtcggggt	ccggcggtt	gggtactcta	gataacctcg	ggccgatcgc	9960
acgcccccg	tggcggcgac	gacccattcg	aacgtctgcc	ctatcaactt	tcgatggtag	10020
tcgcccgtgc	taccatgggt	accacgggtg	acggggaatc	aggggttcgat	tcgggagagg	10080
gagcctgaga	aacgggtacc	acatccaagg	aaggcagcag	gcgcgcaaat	tacccactcc	10140
cgacccgggg	aggtagtgc	gaaaaataac	aatacaggac	tccttcgagg	ccctgtaatt	10200
ggaatgagtc	cacttttaaat	cctttaacga	ggatccattg	gagggcaagt	ctggtgccag	10260
cagccgcggg	aattccagct	ccaatagcgt	atattaaagt	tgctgcagtt	aaaaagctcg	10320
tagttggatc	tggggagcgg	gcgggcggtc	cgcggcgagg	cgagtaccgc	cccgtccccg	10380
cccttgccct	ctcggcgccc	cctcgatgct	cttagctgag	tgtcccgcgg	ggcccgaagc	10440
ggtttactttg	aaaaaattag	agtgttcaaa	gcaggccccga	gcccctgga	taccgcagct	10500
aggaataatg	gaataggacc	gagggtctct	tttgttggtt	ttcggaactg	agggcatgat	10560
taagaggggac	ggccgggggg	attcgtattg	cgcgcgtaga	ggtgaaattc	ttggaccggc	10620

gcaagacgga	ccagagcgaa	agcatttgcc	aagaatgttt	tcattaatca	agaacgaaag	10680
tcggagggttc	gaagacgatc	agataccgtc	gtagttccga	ccataaacga	tgccgactgg	10740
cgatgcgggcg	gcgttatcc	catgacccgc	cgggagcgtt	ccgggaaacc	aaagtctttg	10800
ggttccggggg	ggagtatggg	tgcaaagctg	aaacttaaag	gaattgacgg	aagggcacca	10860
ccaggagtg	gcctgcggt	taatttgact	caacacggga	aacctcacc	ggcccgga	10920
cggacaggat	tgacagattg	atagctcttt	ctcgattccg	tgggtgggtg	tgcatggccg	10980
ttcttagttg	gtggagcgat	ttgtctggtt	aattccgata	acgaacgaga	ctctggcatg	11040
ctaactagtt	acgggacccc	cgagcgggtc	gcgtccccc	acttcttaga	gggacaagt	11100
gcgttcagcc	acccgagatt	gagcaataac	aggtctgtga	tgcccttaga	tgtccggggc	11160
tgacgcgcg	ctacactgac	tggtcagcg	tgtgcctacc	ctgcgcggc	aggcgcgggt	11220
aaccggttga	acccattcg	tgatggggat	cggggattgc	aattattccc	catgaacgag	11280
gaattcccag	taagtgcggg	tcataagctt	gcgttgattt	agtccttggc	ctttgtacac	11340
accgcgcgtc	gctactaccg	attggatggt	ttagtgagge	cctcgatcg	gcccgcgg	11400
ggtcggccca	cggccctggc	ggagcgctga	gaagacggtc	gaacttgact	atctagagga	11460
agtaaaagtc	gtaacaaggt	ttcgttaggt	gaacctgcgg	aaggatcatt	aaacgggaga	11520
ctgtggagga	gcggcgcggt	ggcccgctct	ccccgtcttg	tgtgtgtcct	cgcggggagg	11580
cgcgtgcgtc	acgggacccc	tcgcccgcgt	gtggagcgag	gtgtctggag	tgagggtaga	11640
gaaggggtgg	gtggggctgg	tctgggtccg	tctgggaccg	cctccgattt	cccctcccc	11700
tccctctcc	ctcgtccggc	cttgacctcg	ccaccctacc	gcggcgccgg	ctgctcgcgg	11760
gcgttctgcc	tctttcccg	cgggtctctc	cgtgtctacg	aggggcggt	cgtcgttacg	11820
ggtttttgac	ccgtcccggg	ggcgttcggg	cgtcggggcg	cgcgctttgc	tctcccgga	11880
cccatcccc	ccgcggctct	ggcttttcta	cgttggctgg	ggcggttgtc	gcgtgtgggg	11940
ggatgtgagt	gtcgcgtgtg	ggctcgcccg	acgtttttct	acgtttttct	ggcctcgct	12000
gtcctcccc	ctcctgtccc	gggtacctag	ctgtcgcgtt	ccggcgcgga	ggtttaagga	12060
ccccgggggg	gtcgccttgc	cgccccagg	gtcggggggc	gggtggggccc	gtaggggaagt	12120
ccgtcgttcg	ggcggtctct	cctcagactc	catgaccctc	ctcccccg	tgccgcggt	12180
cggagggcg	cggtcgtgtg	gggggggtga	tgtctggagc	cccctcgggc	gccgtggggg	12240
cccgacccgc	gcccgcgggt	tgcccagattt	cccgcggtcg	gtcctgtcgg	tgccggtcgt	12300
gggttcccg	gtcgttccc	tgtttttccg	ctcccgacc	tttttttttc	ctcccccca	12360
caagtgtctc	gtttcgttcc	tgctggcccg	cctgaggcta	cccctcggtc	catctgttct	12420
cctctctctc	cggggagagg	agggcggtgg	tcttggggg	actgtgccgt	cgtcagcacc	12480
cgtgagttcg	aaatacccg	ctagctgcga	agactctta	gcggtggatc	actcgcctcg	12540
catcgacact	tcgaacgcac	ttgcggccccc	gggttctctc	gaattgcagg	acacattgat	12600
cgtcggttga	cgatcaatcg	cgtcacccgc	tgcggtgggt	cgggggtacg	cctgtctgag	12660
ctcgcagggc	caacccccca	acccgggtcg	ggccctcctg	gctgcgcggc	tgggagtttg	12720
gggcgggttgt	cggtgtggcg	cgcgcgcggc	cgtcgcggag	ctcccgaaagt	tcagacgtgt	12780
cgcgctcgcg	gcttcttccc	gctccgcgt	tcccgccctc	cctgggtctcc	cccgcgcac	12840
gcctcgcgtc	ggcgccctcc	ggaccgctgc	ctcacccagc	ggcggtcctg	cccgtgcccc	12900
tgggaaccca	ccgcgccccc	gtggcgcccg	ggggtggggc	cgtccgcac	tgctctgggt	12960
gaggttggcg	gttgaggggtg	tgctgcgcgc	gaggtgtgtg	tcggtcccct	gcggccgcgg	13020
ggttgtcggg	gtggcggtcg	acgagggccg	gtcgtgcgc	tgccggtggt	gtctgtgtgt	13080
gtttgggtct	tgcgctgggg	gagggcgggg	cgaccgctcg	cggggttggc	gcggtcgccc	13140
ggcgccgcgc	acccctccggc	ttgtgtggag	ggagagcgag	ggcgagaacg	gagagaggtg	13200
gtatcccccg	tgccgttgcg	agggaggggt	tgccgtcccg	cgtccgtccg	tcccctccc	13260
cctcggtggg	cgccttcgcg	ccgcacgcgg	ccgctagggg	cggtcggggc	ccgtggcccc	13320
cgtggctctt	cttcgtctcc	gcttctcctt	caccgcggcg	gtaccgcgtc	cggcgccggc	13380
ccgcgggacg	ccgcggcgctc	cgtgcgcgga	tgcgagtcac	cccgggtgtg	tgcgagttcg	13440
gggagggaga	gggcctcgct	gaccggttgc	gtcccggctt	ccctgggggg	gaccgcggct	13500
ctgtgggtcg	tgccgtcccg	gggttgcgtg	tgagtaagat	cctccacccc	cgcgcgcctc	13560
ccctcccgcc	ggcctctcgg	ggacccccc	agacggttcg	ccggctcgtc	ctcccgtgcc	13620
gcggggtgcc	gtctctttcc	cgcccgccctc	ctcgtctctc	tcttcccgcg	gctggggcg	13680
tgtccccctt	ttctgaccgc	gacctcagat	cagacgtggc	gaccgcgtga	atttaagcat	13740
attagtccgc	ggaggaaaag	aaactaacca	ggattccctc	agtaacggcg	agtgaacagg	13800
gaagagccca	gcgcggaatc	cccgcgcgcg	gtcgcggcgt	gggaaatgtg	gcgtacggaa	13860
gacccactcc	ccggcgccgc	tcgtgggggg	cccaagtcct	tctgatcgag	gcccagcccc	13920
tggacgggtg	gagggccggt	gcggccccc	cgcgcgggc	tcgggtcttc	ccggagtcgg	13980
gttgcttggg	aatgcagccc	aaagcgggtg	gtaaactcca	tctaaggcta	aataccggca	14040
cgagaccgat	agtcaacaag	taccgtaagg	gaaagtgtga	aagaactttg	aagagagagt	14100
tcaagagggc	gtgaaaacgt	taagaggtaa	acgggtgggg	tcgcgcgag	ccgcccggag	14160
gattcaaccc	ggcggcgcgc	gtccggccgt	gcccgggtgt	cccggcgga	ctttcccgct	14220
ccccgttctt	cccgaccctt	ccaccgcgcg	gtcgttcccc	tcttctctcc	cgcgtccggc	14280
gcctccggcg	gcgggcgcgg	gggggtgggt	gggtgtggcg	cgcggggcg	gcgggggtg	14340
gggtcggcg	gggaccgccc	ccggccggcg	accggccgcc	gccgggcgca	cttccaccgt	14400
ggcggtgcgc	cgcgaccggc	tccgggacgg	ccgggaaagc	ccgggtggga	aggtggctcg	14460
ggggggcgcg	cgcgtctcag	ggcgcccgga	accacctcac	cccgagtgtt	acagccctcc	14520
ggcgcgctt	tcgccgaatc	ccggggccga	ggaagccaga	taccgctcgc	cgcgctctcc	14580

ctctcccccc	gtccgcctcc	cgggcgggcg	tgggggtggg	ggccggggcg	cccctccac	14700
ggcgcgaccc	ctctccacc	cccctccgtc	gcctctctcg	gggcccgggtg	ggggcggggg	14760
cggactgtcc	ccagtgcgcc	ccgggctgcg	tcgcgcgcgtc	gggtcccggg	gggaccgtcg	14820
gtcacgcgtc	tcccgaacga	gcccagcgca	cggggtcggc	ggcgatgtcg	gctacccacc	14880
cgacccgtct	tgaaacacgg	accaaggagt	ctaaccgcgtg	cgcgagtcag	gggctcgtcc	14940
gaaagccgcc	gtggcgcaat	gaaggtgaag	ggccccgcgc	ggggggccga	ggtgggatcc	15000
cgaggcctct	ccagtcgcc	gagggcgcac	caccggcccg	tctcgcccg	cgcgccgggg	15060
aggtggagca	cgagcgtacg	cggttaggacc	cgaaagatgg	tgaactatgc	ttgggcaggg	15120
cgaagccaga	ggaaactctg	gtggagggtcc	gtagcggtcc	tgacgtgcaa	atcggtcgtc	15180
cgacctgggt	atagggcgga	aagactaatc	gaacctcta	gtagctggtt	ccctccgaag	15240
tttccctcag	gatagctggc	gctctcgctc	ccgacgtacg	cagttttatc	cggtaaagcg	15300
aatgattaga	ggtcttgggg	ccgaaacgat	ctcaacctat	tctcaaaactt	taaaatgggta	15360
agaagcccg	ctcgctggcg	tggagccggg	cggtgaatgc	gagtgcctag	tgggccaactt	15420
ttggttaagca	gaactggcg	tgcgggatga	accgaacgcc	gggttaaggc	gcccgatgcc	15480
gacgctcatc	agaccccaga	aaaggtgttg	ggtgatatag	acagcaggac	ggtggccatg	15540
gaagtccgaa	tccgctaagg	agtgtgtaac	aactcacctg	ccgaatcaac	tagccctgaa	15600
aatggatggc	gctggagcgt	cgggcccata	cccggccgtc	gcccagtcg	gaacggaaacg	15660
ggacggggagc	ggccgcgggt	gcgcgtctct	cggggtcggg	ggtgcgtggc	gggggcccgt	15720
cccccgccctc	ccctccgcgc	gcccgggttcg	cccccgccgc	gtcgggcccc	gcccagccta	15780
cgccgcgacg	agtagggagg	ccgctggcgtt	gagccttgaa	gcctaggggcg	cgggcccggg	15840
tggagccgcc	gcagggtgcag	atcttgggtgg	tagtagcaaa	tattcaaacg	agaacttttga	15900
agggccgaagt	ggagaagggt	tccatgtgaa	cagcagttga	acatgggtca	gtcggtcctg	15960
agagatgggc	gagtgccggt	ccgaaggagc	cgccgttgcc	ctccgttgcc	ctcgcccgat	16020
cgaaagggag	tcgggttcag	atccccgaat	ccggagtggc	ggagatgggc	gcccgcaggc	16080
cagtgcggta	acgcgaccga	tcccggagaa	gcccggcgga	ggcctcgggg	agagttctct	16140
agagatgggc	gagtgccggt	cgccctggaa	gggctcgcc	ccgagagagg	ggcccggtcc	16200
ttggaagcg	tcgcggttcc	ggcgcggtcc	ggtgagctct	cgctggccct	tgaaaatccg	16260
ggggagaggg	tgtaaatctc	gcgcggggcc	gtacccatat	ccgcagcagg	tctccaaggt	16320
gaacagcctc	tggcatgttg	gaacaatgta	ggttaaggaa	gtcggcaagc	cggatccgta	16380
acttcgggat	aaggattggc	tctaagggtc	gggtcggtcg	ggctggggcg	cgaagcgggg	16440
ctgggcgcgc	gcccgcggctg	gacgagggcg	cgccgcctc	tcccacgtcc	ggggagaccc	16500
cccgctcttt	ccgcgcgggc	ccgcccctcc	ctctcccccg	cgggggccccg	tcgtcccccg	16560
cgctcgtegc	acctctcttc	ccccctcctt	cttcccgtcg	ggggcggggt	cggggggtcgg	16620
cgcgcgccgc	gggctccggg	gcggcggggtc	caaccccgcg	gggggttccgg	agcgggagga	16680
accagcggtc	cccgggtggg	cgggggggccc	ggacactcgg	ggggccggcg	gcccgcggca	16740
ctctggacgc	gagccggggc	cttcccgtgg	atcgccctag	ctgcggcggg	cgtcgcggcc	16800
gctcccgggg	agcccggggg	gtgcccggcg	gggtcccctc	cccgcggggc	ctcgctccac	16860
ccccctctcg	cctctcccga	ggtgcgtggc	ggggcggtcc	gggcgtgtcc	cgcgctgtg	16920
gggggaacct	ccgcgtcggt	gttcccccg	cggttcggcc	cccggggccg	cggttttccg	16980
cgcgccgccc	ccgctccggc	cgggcgctag	cagccgactt	agaactgggtg	cggaccaggg	17040
gaatccgact	gtttaattaa	aacaaagcat	cgcgagggcc	cgcgccgggt	gttgacgcga	17100
tgtgattttct	gcccagtgct	ctgaatgtca	aagtgaagaa	attcaatgaa	gcgcgggttaa	17160
acggcgggag	taactatgac	tctcttaagg	tgcccaaatg	cctcgatcatc	taattagtga	17220
cgcgcatgaa	tggatgaacg	agattcccac	tgtccctacc	tactatccag	cgaaaccaca	17280
gccaaaggaa	cgggcttggc	ggaatcagcg	gggaaagaag	accctgttga	gcttgactct	17340
agtcctggcac	ggtgaagaga	catgagaggt	gtagaataag	tgggaggccc	ccggcgcccg	17400
gcccctcctc	cgctcgggg	tcggggcaccg	ccggcctcgc	gggcccggcg	tgaaatacca	17460
ctactctcat	cgttttttca	ctgacccgggt	gaggcggggg	ggcgagcccc	gaggggctct	17520
cgcttctggc	gccaagcgtc	cgctcccgcgc	gtgcccggcg	gcgcgacccg	ctccggggac	17580
agtgccagggt	ggggaggttg	actggggcgg	tacactcgctc	aaacggtaac	cagggtgtcc	17640
taaggcgagc	tcaggggagga	cagaaacctc	ccgtggagca	gaaggggcaaa	agctcgcttg	17700
atcttgatttt	tcagtaacga	tacagaccgt	gaaagcgggg	cctcacgatc	cttctgacct	17760
tttgggtttt	aagcaggagg	tgtcagaaaa	gttaccacag	ggataactgg	cttgtggcgg	17820
ccaagcgttc	atagcgacgt	cgcttttttga	tccctcgatg	tcggctcttc	ctatcattgt	17880
gaagcagaat	tcaccaagcg	ttggattggt	caccactata	tagggaaacgt	gagctgggtt	17940
tagaccgtcg	tgagacagggt	tagttttacc	ctactgatga	tgtgttgttg	ccatggtaat	18000
cctgctcagt	acgagaggaa	ccgcaggttc	agacatttgg	tgtatgtgct	tggctgagga	18060
gccaatgggg	cgaagctacc	atctgtggga	ttatgactga	acgcctctaa	gtcagaatcc	18120
gcccaagcgg	aacgatacgg	cagcgccgaa	ggagcctcgg	ttggcccccg	atagccgggt	18180
ccccgtccgt	cccgctcggc	gggggtccccg	cgtcgccccg	cgggcgccgc	gggtctcccc	18240
ccgcggggcg	tcgggacggg	ggtccgggtgc	ggagagccgt	tcgtcttggg	aaacgggggtg	18300
cgcccggaata	gggggcccgc	ctctcgcccg	tcacgttgaa	cgacgttcg	tgtggaacct	18360
ggcgctaaac	cattcgtaga	cgacctgctt	ctgggtcggg	gtttcgtagc	tagcagagca	18420
gctccctcgc	tgcgacttat	tgaaagtcag	ccctcgacac	aagggtttgt	ctctcggggc	18480
tttcccgctg	cacgcccgtc	cgctcgacg	cgacgtgtc	gcccggcggg	cgctcacgggg	18540
cggttcgctc	gcccggcg	gcggttgccc	gaacgacgt	gtgggtgggtg	ggggggggat	18600
cgtcttctcc	tccgtctccc	gaggacgggtt	cgtttctctt	tccccttccg	tcgctctcct	18660

tgggtgtggg	agcctcgtgc	cgtcgcgacc	gcgccctgcc	gtcgcctgcc	gccgcagccc	18720
cttgccctcc	ggccttggcc	aagccggagg	gcgaggagg	gggatcggcg	gccgcggcga	18780
ccgcggcgcg	gtgacgcacg	gtgggatccc	catcctcggc	gcgtccgtcg	gggacggccg	18840
ggtggagggg	cgggaggggt	ttttcccggt	aacgcgcggt	tcggcgccag	gcctctggcg	18900
gccggggggg	cgctctctcc	gccccagcat	ccccactccc	gccccctctc	ttcgcgcgcc	18960
gcggcggcga	cgtgcgtacg	aggggaggat	gtcgcgggtg	ggaggcggag	agggtccggc	19020
gcggcgcctc	ttccattttt	tcccccccaa	cttcggagggt	cgaccagtac	tcggggcgac	19080
actttgtttt	ttttttttcc	cccgatgctg	gaggtcgacc	agatgtccga	aagtgtcccc	19140
ccccccccc	ccccccggcg	cggagcggcg	gggccactct	ggactctttt	tttttttttt	19200
tttttttttt	ttaaattcct	ggaaccttta	ggtcgaccag	ttgtccgtct	tttactcctt	19260
catataggtc	gaccagtact	ccgggtggta	ctttgtcttt	ttctgaaaaa	cccagagggtc	19320
gaccagatat	ccgaaagtcc	tctctttccc	tttactcttc	cccacagcga	ttctcttttt	19380
tttttttttt	tttgggtgtg	ctcttttttg	cttatataca	tgtaaatagt	gtgtacgttt	19440
atatacttat	aggaggagg	cgaccagtac	tcggggcgac	actttgtttt	tttttttttt	19500
tcaccagatg	atggagggtcg	accagatgtc	cgaaagtgtc	ccgtccccc	cctccccc	19560
ccgcgacgcg	gcggggtcac	tctggactct	tttttttttt	tttttttttt	tttaaatttc	19620
tggaaacctta	aggtcgacca	gttgtccgtg	tttcaactcat	tcataatagg	cgaccgggtg	19680
taactttgtct	ttttctgaaa	atcgagagg	tcgaccagat	gtcagaaagt	ctggtgtgtcg	19740
ataaattatc	tgatctagat	ttgtttttct	gttttttcagt	tttgtgttgt	tttgtgttgt	19800
tttgtgttgt	tttgttttgt	tttgttttgt	tttgttttgt	tttgttttgt	tttgttttgt	19860
tttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	19920
gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	gttgtgttgt	19980
ttgtttgtctg	ttgtttgtgt	ttgttcgggt	cgaacagttg	tccttaaccg	agtttttttg	20040
tacacaaaca	tgcacttttt	ttaaaataaa	tttttaaaat	aaatgcgaaa	atcgaccaat	20100
tatccctttc	cttctctctc	tttttaaaaa	attttctttg	tgtgtgtgtg	tgtgtgtgtg	20160
tgtgtgtgtg	tcggtgtgtg	tgtgtgtgtg	cgtgcagcgt	gcgcgcgcctc	gttttataaa	20220
taactataat	aataggtcgc	cgggtggttg	tagcttcccg	gactccagag	gcagaggcag	20280
gcagacttct	gagttcgagg	ccagcctgtg	ctacagagga	accctgtctc	gaaaaatgaa	20340
aataaataca	tacatacata	catatacata	catatacata	catatacata	tacatatgag	20400
gttgaccagt	tgtcaatcct	ttagaatttt	gtttttaatt	aatgtgatag	agagatagat	20460
aatagataga	tggatagagt	gatacaaaat	taggtttttt	tttcagtaaa	tatgagggtg	20520
attaaccact	tttccctttt	taggtttttt	ttttttccc	ctgtccatgt	ggttgctggg	20580
atttgaactc	aggaccctgg	caggtcactt	ggaaaacgtg	ttttctatat	atataaatag	20640
tgggtctgtct	gctgttttgt	tgtttgcttg	cttgcttgct	tgcttgcttg	cttgcttgct	20700
tgtttttttt	tttctcttga	gacagtattt	cctgtgttaa	cctgggtgcc	tgaaactcac	20760
tctgtagacc	agcctggcct	caatcgaaact	cagaaatcct	cctgcctctt	gtctacctcc	20820
caatttttga	gtaaagggtg	gctacaccac	tgcttggcat	tattatcatt	atcattatta	20880
atattttatt	tagacagaac	gaaatcaact	agttgggtct	gtttcgttaa	ttcatttgaa	20940
attagttaga	ccaattagtt	ggctggtttg	ggaggtttct	tttggttccg	atttgggtgt	21000
ttgtggggct	ggggatcagg	tatctcaacg	gaatgcatga	agggttaagg	gagatggctc	21060
gattttttga	aagattactt	ttcttagtct	gaggaaaaaa	taaaataata	ttgggctacg	21120
tttcaattgt	tcattttctat	ttctttctct	ttctttcttt	ctttcagata	aggaggtcgg	21180
ccagttctct	ctgccttctg	gaagatgtag	gcattgcatt	gggaaaagca	ttgtttgaga	21240
gatgtgctag	tgaaccagag	agttttggatg	tcaagccgta	taatgtttat	tacaatatag	21300
aaaagttcta	acaaagtgat	ctttaacttt	tttttttttt	tttctccttc	tacttctact	21360
tgttctcact	ctgccacca	cgcgctttgt	acattgaaatg	tgagctttgt	tttgcttaac	21420
agacatatat	tttttctttt	ggtttttgctt	gacatggttt	ccctttctat	ccgtgcaggg	21480
ttcccagacg	gccttttgag	aataaaaatgg	gaggccagaa	ccaaagtctt	ttgaataaag	21540
caccacaact	ctaacctgtt	tggctgtttt	ccttcccaag	gcacagatct	ttcccagcat	21600
ggaaaagcat	gtagcagttg	taggacacac	tagacgagag	caccagatct	cattgtgggt	21660
ggttgtgaac	cacccaccat	gtggttgctt	gggatttgaa	ctcaggatct	tcagaagacg	21720
agtcagggct	ctaaaccgat	gagccatctc	tccagccctc	ctacattcct	tcttaaggca	21780
tgaatgatcc	cagcatggga	agacagtctg	ccctctttgt	ggtatatcac	catatactca	21840
ataaaaataat	gaaatgaatg	aagtctccac	gtatttattt	cttcgagcta	tctaaattct	21900
ctcacagcac	ctccccctcc	cccacactgc	ctttctccct	atgtttgggt	ggggctgggg	21960
gaggggtggg	gtggggggcag	ggatctgcat	gtcttctttg	aggctctgtga	actatttgcg	22020
atggcctggg	tctctgaact	gttgagcctt	gtctatccag	aggctgactg	gctagttttc	22080
tacctgaagt	ccctgagtga	tgattttccct	gtgaattc			22118

<210> 19  
 <211> 175  
 <212> DNA  
 <213> Mus musculus

<400> 19	ctcccgccg	gcccccggtg	tcgcccgttc	cgtggcgccg	acaatgcggt	tgtgcgtcca	60
cgtgtgcgtg	tccgtgcagt	gccgttgtgg	agtgccctgc	tctcctctc	ctccccggca		120



gcgttccac ggttggggac caccggtgac ctgcctct tcgggcctgg atccg

175

<210> 20  
<211> 755  
<212> DNA  
<213> Mus musculus

<400> 20							
ggtctggtg	gaattgttga	cctcgctctc	gggtgcggcc	tttggggaac	ggcggggtcg		60
gtcgtgcccg	gcgcgggacg	tgtgtcgggg	cccacttccc	gctcgagggg	ggcgggtggcg		120
gcggcggttg	tagtctcccg	tgttgcgtct	tcccgggctc	ttgggggggg	tgccgtcggt		180
ttcggggccg	gcgttgcttg	gcttacgcag	gcttggtttg	ggactgcctc	aggagtcgtg		240
ggcgggtgtga	ttcccgcggg	ttttgcctcg	cgtctgcctg	ctttgcctcg	ggtttgcttg		300
gttcgtgtct	cgggagcggt	ggtttttttt	tttttcgggt	cccggggaga	ggggtttttc		360
cgggggacgt	tcccgtcgcc	ccctgcgcgc	ggtgggtttt	cgtttcgggc	tgtgttcgtt		420
tccccttccc	cgtttcgccg	tcggttctcc	cgggtcggtc	ggccctctcc	ccggtcggtc		480
gcccggccgt	gctgccggac	cccccttctt	gggggggatg	cccggggcacg	cacgcgtccg		540
ggcggccact	gtggtccggg	agctgctcgg	caggcgggtg	agccagttag	aggggctca		600
tgcccccgcg	ggctcccgtg	gccgacgcgg	cgtgttcttt	ggggggggcct	gtcgtgcgg		660
gaaggctcgc	cacgttgtcg	gtccttgcca	gggaaagagg	cttttttttt	ttagggggtc		720
gtccttcgtc	gtcccgtcgg	cggtggatcc	ggcct				755

<210> 21  
<211> 463  
<212> DNA  
<213> Mus musculus

<400> 21							
ggccgaggtg	cgtctgcggg	ttggggctcg	tcgggccccg	tcgtcctccg	ggaaggcgtt		60
tagcgggtac	cgtcgcccg	ccgaggtggg	cgcacgtcgg	tgagataaac	ccgagcgtgt		120
ttctggttgt	tggcggcggg	ggctccggtc	gatgtcttcc	cctccccctc	tccccgagge		180
caggtcagcc	tcgcctctgt	ggcttcgtcg	gccgtctccc	ccccctcac	gtccctcggc		240
agcgagcccg	tccgttcgac	cttccttccg	ccttcccccc	atctttccgc	gctccgttgg		300
ccccgggggt	ttcacggcgc	ccccacgct	cctccgcctc	tcgcgccgtg	gtttggacgc		360
ctggttcggg	tctccccgcc	aaaccccggt	tgggttggtc	tcgggccccg	gcttgctctt		420
cgggtctccc	aacccccggc	cggaagggtt	cgggggttcc	ggg			463

<210> 22  
<211> 378  
<212> DNA  
<213> Mus musculus

<400> 22							
ggattcttca	ggattgaaac	ccaaaccggt	tcagtttctt	ttccggctcc	ggccgggggg		60
ggcggccccc	ggcgggtttg	tgagttagat	aacctcgggc	cgatcgcacg	ccccccgtgg		120
cggcgacgac	ccattcgaac	gtctgcccta	tcaactttcg	atggtagtcg	atgtgcctac		180
catggtgacc	acgggtgacg	gggaatcagg	gttcgattcc	ggagagggag	cctgagaaac		240
ggctaccaca	tccaaggaag	gcagcaggcg	cgcaaattac	ccactcccga	cccggggagg		300
tagtgacgaa	aaataacaat	acaggactct	ttcgaggccc	tgtaattgga	atgagtcac		360
tttaaatcct	ttaagcag						378

<210> 23  
<211> 378  
<212> DNA  
<213> Mus musculus

<400> 23							
gatccatttg	agggcaagtc	tgggtgccagc	agccgcggta	attccagctc	caatagcgtg		60
tattaaagtt	gctgcagtta	aaaagctcgt	agttggatct	tgggagcggg	cgggcgggtcc		120
gccgcgaggg	gagtcaccgc	ccgtccccgc	cccttgccctc	tcggcgcccc	ctcgatgctc		180
ttagctgagt	tgtcccgcgg	ggcccgaagc	gtttactttg	aaaaaattag	agttgtttca		240
aagcaggccc	gagccgcctg	gataccgcca	gctaggaaat	aatggaatag	gaccgcgggt		300
cctattttgt	ttggttttcg	gaactgagcc	catgattaag	ggaaacggcc	gggggcatte		360
ccttattgcg	ccccctta						378

<210> 24  
<211> 719

<212> DNA

<213> Mus musculus

<400> 24

ggatctttcc	cgctccccgt	tcttccccgg	ccctccaccc	gagcgctctcc	ccccctcttt	60
tccccctctcc	ggaggggggg	gaggtggggg	cgcggtgggg	gggtcggggg	tggggtcggg	120
ggggggaccgc	ccccggccgg	caaaaaggccg	ccgcggggcg	cacttcaacc	gtagcggtgc	180
gcccgcgaccg	gctacgagac	ggctgggaag	gcccgcaggg	gaatgtggct	cgggggggggc	240
ggcgcgctctc	agggcgcgcc	gaaccacctc	accccgagtg	ttacagccct	ccggcccgccg	300
tttcgcggaa	tccccggggc	gaggggaagc	cgatacccc	tcgcgcgcgt	tttccccctcc	360
ccccgtccgc	ctcccggggc	ggcggtgggg	tggggggccg	gccgcccctc	ccacgcccgt	420
ggtttctctc	tctcccgggt	tcggccgggt	tggggggggg	agcccgggtg	ggggcggggg	480
ggactgtcct	cagtgcgccc	cgggcgctcg	cgcgccgctg	ggcccggggg	gttctctcgg	540
tcacgcggcc	cccgcggaag	ccgagcgcac	ggggtcggcg	gcgatgtcgg	ctaccacccc	600
gacccgtctt	gaaacacgga	ccaaggagtc	taacgcgtgc	gcgagtcagg	ggctcgcacg	660
aaagccgcgg	tggcgcaatg	aaggtgaagg	gccccgtccg	ggggcccgag	gtgggatcc	719

<210> 25

<211> 685

<212> DNA

<213> Mus musculus

<400> 25

cgaggcctct	ccagtccgcc	gagggcgcac	caccggcccg	tctcgcccg	cgcgtcgggg	60
aggtggagca	cgagcgtacg	cgtaggacc	cgaaagatgg	tgaactatgc	ctgggcaggg	120
cgaagccaga	ggaaactctg	gtggaggtcc	gtagcggtcc	tgacgtgcaa	atcggtcgtc	180
cgacctgggt	ataggggcga	aagactaatc	gaaccatcta	gtagctgggt	ccctccgaag	240
tttccctcag	gatagctggc	gctctcgcaa	ccttcggaag	cagttttatc	cggttaaagg	300
cggaaatggat	taggaggtct	tggggccgga	aacgatctca	aactatttct	caaactttaa	360
atgggtaagg	aagcccggct	cgctggcggt	gagccggggc	tggaaatgca	gtgcctagt	420
ggccactttt	ggttaagcaga	actggcgctg	cgggatgaac	cgaacgcggg	gttaaggcgc	480
ccgatccga	cgctcatcag	accccagaaa	aggtgttgg	tgatatagac	agcaggacgg	540
tggccatgga	agtcggaatc	cgctaaggag	tgtgtaacaa	ctcacctgcc	gaatcaacta	600
gccctgaaaa	tggatggcgc	tggagcgtcg	ggcccatacc	cgggcgctcg	cggcagtcgg	660
aacgggacgg	gacgggagcg	gcccgc				685

<210> 26

<211> 5162

<212> DNA

<213> Artificial Sequence

<220>

<223> Chimeric bacterial plasmid

<400> 26

gacggatcgg	gagatctccc	gatccccctat	ggtcgactct	cagtacaatc	tgctctgatg	60
ccgcatagtt	aagccagtat	ctgctccctg	cttgtgtgtt	ggaggtcgct	gagttagtgc	120
cgagcaaaat	ttaagctaca	acaaggcaag	gcttgaccga	caattgcatg	aagaatctgc	180
ttaggggttag	gcgtttttgcg	ctgcttcgcg	atgtacgggc	cagatatacg	cgttgacatt	240
gattattgac	tagttattaa	tagtaatcaa	ttacgggggtc	attagttcat	agcccatata	300
tggagttccg	cgttacataa	cttacggtaa	atggcccgcg	tggctgaccc	cccaacgacc	360
ccgcgccatt	gacgtcaata	atgacgtatg	ttcccatagt	aacgcccaata	gggactttcc	420
attgacgtca	atgggtggac	tatttacggg	aaactgcccc	cttggcagta	catcaagtgt	480
atcatatgcc	aagtaacgcc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	540
atgcccagta	catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	600
tcgctattac	catggtgatg	cggtttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	660
actcacgggg	atttccaagt	ctocacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	720
aaaatcaacg	ggactttcca	aaatgtcgta	acaactccgc	cccatgacg	caaattggcg	780
gtaggcggtg	acgggtgggag	gtctatataa	gcagagctct	ctgggctaact	agagaaccca	840
ctgcttactg	gcttatcgaa	attaatacga	ctcactatag	ggagacccaa	gcttggtaac	900
gagctcggat	cgatatctgc	ggccgcgctg	acgggaattca	gtggatccac	tagtaacggc	960
cgccagtgtg	ctggaattaa	ttcgctgtct	gcgaggggcca	gctgttgggg	tgagtactcc	1020
ctctcaaaag	cgggcatagc	ttctgcgcta	agattgtcag	tttccaaaaa	cgaggaggat	1080
ttgatattca	cctggcccg	ggtgatgcct	ttgaggggtg	ccgcgtccat	ctgggcagaa	1140
aagacaatct	ttttgttgtc	aagcttgagg	tgtggcaggc	ttgagatctg	gccatacact	1200
tgagtgaaca	tgacatccac	tttgcccttc	tctccacagg	tgtccactcc	caggtccaac	1260
tgcaggtcga	gcattgcatt	agggcgggcca	attccgcccc	tctccctccc	ccccccctaa	1320

cgttactggc	cgaagccgct	tggaataagg	cgggtgtgcg	tttgtctata	tgtgattttc	1380
caccatattg	ccgtcttttt	gcaatgtgag	ggcccggaaa	cctggccctg	tcttcttgac	1440
gagcattcct	aggggtcttt	cccctctcgc	caaagggaatg	caaggtctgt	tgaatgtcgt	1500
gaaggaaagca	gttcctctgg	aagcttcttg	aagacaaaaca	acgtctgtag	cgaccctttg	1560
caggcagcgg	aaacccccac	ctggcgacag	gtgcctctgc	ggccaaaagc	cacgtgtata	1620
agatacacct	gcaaaggcgg	cacaacccca	gtgccacgtt	gtgagttgga	tagttgtgga	1680
aagagtcaaa	tggctctcct	caagcgtatt	caacaagggg	ctgaaggatg	cccagaaggt	1740
accccatgtg	atgggatctg	atctggggcc	tgggtgcaca	tgctttacat	gtgttttagtc	1800
gagggtaaaa	aaacgtctag	gccccccgaa	ccacggggac	gtggttttcc	tttgaaaaaac	1860
acgatgataa	gcttgccaca	acccggggtc	caccggctgc	caccatgggtg	agcaagggcg	1920
aggagctggt	caccgggggtg	gtgcccattcc	tggctcagct	ggacggcgac	gtaaacgggc	1980
acaagttcag	cgtgtccggc	gagggcgagg	gcgatgccac	ctacggcaag	ctgaccctga	2040
agttcatctg	caccaccggc	aagctgcccc	tgccctggcc	caccctcgtg	accaccctga	2100
cctacggcgt	gcagtgcttc	agccgctacc	ccgaccacat	gaagcagcac	gacttcttca	2160
agtccgccat	gcccgaaggc	tacgtccagg	agcgcaccat	cttcttcaag	gacgacggca	2220
actacaagac	ccgcgccgag	gtgaagttcg	agggcgacac	cctggtgaac	cgcacgcagc	2280
tgaagggcat	gcacttcaag	gaggacggca	acatcctggg	gcacaagctg	gagtacaact	2340
acaacagcca	caacgtctat	atcatggccg	acaagcagaa	gaacggcatc	aaggtgaact	2400
tcaagatccg	ccacaacatc	gaggacggca	gcgtgcagct	cgccgaccac	taccagcaga	2460
acacccccat	cggcgacggc	cccgtgctgc	tgcccagcaa	ccactacctg	agcaccagat	2520
ccgccttgag	caaagacccc	aacgagaagc	gcgatcacat	ggtcctgctg	gagttcgtga	2580
ccgcccgcgg	gatcactctc	ggcatggacg	agctgtacaa	gtaaagcggc	cctagagctc	2640
gctgatcagc	ctcgactgtg	cctctagtgt	ccagccatct	gttggtttgcc	cctccccgtg	2700
gccttccttg	accctggaag	gtgccactcc	cactgtcctt	tcctaataaaa	atgaggaaat	2760
tgcctcgcat	tgtctgagta	ggtgtcattc	tattctgggg	ggtgggggtg	ggcaggacag	2820
caagggggag	gattggggaag	acaatagcag	gcatgctggg	gatgcgggtg	gctctatggc	2880
ttctgaggcg	gaaagaacca	gctggggctc	gagtgcatte	tagttgtggg	ttgtccaaac	2940
tcacatcaat	atctttatcat	gtctgtatac	cgtcgacctc	tagctagagc	ttggcgtaat	3000
catggcataa	gctgttctct	gtgtgaaatt	gttatccgct	cacaattcca	cacacatac	3060
gagccggaag	cataaagtgt	aaagcctggg	gtgcctaatt	agtgcgctaa	ctcacattaa	3120
ttgcgttgcc	ctcactgccc	gctttccagt	cgggaaaacct	gtcgtgccag	ctgcattaat	3180
gaatcgagat	acgcgcgggg	agaggcggtt	tgcgatttgg	gcgctcttcc	gcttccctcg	3240
tcactgactc	gctgcgctcg	gtcggttcggc	tgccggcgagc	ggtatcagct	cactcaaagg	3300
cggtaataacg	gttatccaca	gaatcagggg	ataacgcgag	aaagaacatg	tgagcaaaag	3360
gcccacaaa	ggccaggaaac	cgtaaaaagg	ccgcgttgct	ggcgtttttc	cataggctcc	3420
gccccctga	cgagcatcac	aaaaaatcgag	gctcaagtca	gaggtggcga	aaccgcagag	3480
gactataaag	ataccaggcg	tttccccctg	gaagctccct	cgtgcgctct	cctgttccga	3540
ccctgcgcgt	taccggatag	ctgtccgcct	ttctcccttc	gggaagcgtg	gcgctttctc	3600
aatgctcacg	ctgtaggtat	ctcagttcgg	tgtaggctcg	tcgctccaag	ctgggctgtg	3660
tgcacgaacc	ccccgttcag	cccgaaccgt	gcgccttatc	cggtaactat	cgtcttgagt	3720
ccaaccgggt	aagacacgac	ttatcgccac	tggcagcagc	cactggtaac	aggattagca	3780
gagcgaggta	gttaggggtg	gctacagagt	tcttgaaagt	gtggcctaac	tacggctaca	3840
ctagaaggac	agtatttggt	atctgcgctc	tgctgaagcc	agttaccttc	ggaaaaagag	3900
ttggtagctc	ttgatccggc	aaacaaaacca	ccgctggtag	cgggtggtttt	tttgtttgca	3960
agcagcagat	tacgcgcaga	aaaaaaggat	ctcaagaaga	tcctttgatc	ttttctacgg	4020
gggtctgacgc	tcagtggaaac	gaaaactcac	gttaagggat	tttggtcatg	agattatcaa	4080
aaaggatctt	cacctagatc	cttttaaat	aaaaatgaag	ttttaaatca	atctaaaagta	4140
tatatgagta	aacttgggtc	gacagttacc	aatgcttaat	cagtgaaggca	cctatctcag	4200
cgatctgtct	atttcggttca	tccatagttg	cctgactccc	cgtcgtgtag	ataactacga	4260
tacgggaggg	cttaccatct	ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	4320
cggctccaga	tttatcagca	ataaaccagc	cagccggaag	ggccgagcgc	agaagtggtc	4380
ctgcaacttt	atccgcctcc	atccagtcta	ttaattgttg	ccgggaagct	agagtaagta	4440
gttcgccagt	taatagtttg	cgcaacgttg	ttgccattgc	tacaggcatc	gtgggtgtcac	4500
gctcgtcggt	tggtatggct	tcattcagct	ccggttccca	acgatcaagg	cgagttacat	4560
gatcccccat	gttgtgcaaa	aaagcgggta	gctccttcgg	tcctccgac	gttgtcagaa	4620
gtaagttggc	cgcagtgtta	tcactcatgg	ttatggcagc	actgcataat	tcctctactg	4680
tcatgccatc	cgtaagatgc	ttttctgtga	ctgggtgagta	ctcaaccaag	tcattctgag	4740
aatagtgtat	gcggcgaccc	agttgctctt	gcccggcgct	aatacgggat	aataccgcgc	4800
cacatagcag	aactttaaaa	gtgctcatca	ttggaaaaac	ttcttcgggg	cgaaaaactct	4860
caaggatctt	accgctgttg	agatccagtt	cgatgtaacc	cactcgtgca	cccaactgat	4920
cttcagcatc	ttttactttc	accagcggtt	ctgggtgagc	aaaaacagga	aggcaaaatg	4980
ccgcaaaaaa	gggaataaagg	gcgacacggg	aatgttgaat	actcatactc	ttcctttttc	5040
aatattattg	aagcatttat	cagggttatt	gtctcatgag	cggatacata	tttgaatgta	5100
tttagaaaaa	taaacaaata	gggggttccgc	gcacatttcc	ccgaaaagtg	ccacctgacg	5160
tc						5162

<211> 5627  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> pMG plasmid from InvivoGen; IRES sequence modified  
EMCV nucleotides 2736-3308

<400> 27  
caccggcgaa ggaggcctag atctatcgat tgtacagcta gctcgacatg ataagataca 60  
ttgatgagtt tggacaaacc acaactagaa tgcagtgaaa aaaatgcttt atttgtgaaa 120  
tttgtgatgc tattgcttta tttgtgaaat ttgtgatgct attgctttat ttgtaaccat 180  
tataagctgc aataaacaag ttaacaacaa caattgcatt cattttatgt ttcaggttca 240  
gggggagggtg tgggagggtt tttaaagcaa gtaaaacctc tacaaatgtg gtagatccat 300  
ttaaatgtta attaagaaca tgtgagcaaa agggccagca aaggccagga accgtaaaaa 360  
ggccgcggtg ctggcggttt tccataggct ccgccccctt gacgagcatc acaaaaaatcg 420  
acgtcaagtc cagagctggc gaaacccgac aggactataa agataaccagg cgtttcccc 480  
tggaagctcc ctccgtgcgt ctccgtgttc gaccctgccc ettaccggat acctgtccgc 540  
ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggt atctcagttc 600  
gggtgtaggt gttcgctcca agctgggctg tgtgcacgaa ccccccgttc agcccgaccg 660  
ctgcccctta tccggtaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc 720  
actggcagca gccactggta acaggattag cagagcgagg tatgtagggc gtgctacaga 780  
gttcttgaag tggtagccta actacggcta cactagaaga acagtatttg gtatctgcgc 840  
tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaaac 900  
caccgctggt agcgggtggt tttttgtttg caagcagcag attacgcgca gaaaaaaagg 960  
atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtgga acgaaaactc 1020  
acgttaaggg atttttgtca tggctagtta attaagctgc aataaacaat cattattttc 1080  
attggatctg tgtgttggtt ttttgtgtgg gcttggggga gggggaggcc agaattgactc 1140  
caagagctac aggaaggcag gtccagagacc ccactggaca aacagtggct ggactctgca 1200  
ccataacaca caatcaacag gggagtgcgc tagatccgt tagagtccgt tacataactt 1260  
acggtaaatg gccgcgctgg ctgaccgccc aacgaccccc gcccattgac gtcaataatg 1320  
acgtatgttc ccatagtaac gccaataggg actttccatt gacgtcaatg ggtggagtat 1380  
ttacggtaaa ctgcccactt ggcagtagat caagtgtatc atatgccaa gacgtccctc 1440  
attgacgtca atgacgttaa atggcccgcc tggcattatg cccagtacat gaccttatgg 1500  
gactttccta ctgggcagta catctacgta ctattaccat ggtgatgcgc 1560  
ttttggcagt acatcaatgg gcgtggatag cggtttgact caccggggatt tccaagtctc 1620  
caccocatgt acgtcaatgg gagtttgttt tggcaccaaa atcaacggga ctttccaaaa 1680  
tgtcgtaaca actccgccc attgacgcaa atggggcgta ggcgtgtacg gtgggaggtc 1740  
tatataagca gagctcggtt agtgaaccgt cagatcgctt ggagacgcca tccacgtgt 1800  
tttgacctcc atagaagaca ccgggaccga tccagcctcc gcggccggga acggtgcatt 1860  
ggaacgcgga tccccgtgc caagagtgc gtaagatccg cctatagagt ctataggccc 1920  
acccccctgg cttcttatgc atgctatagc gtttttggct tggggtctat acacccccgc 1980  
ttcctcatgt tataggtgat ggtatagctt agcctatagg tgtgggttat tgaccattat 2040  
tgaccactcc cctattgggt acgatacttt ccattactaa cactgtcct tccataacat ggctctttgc 2100  
cacaactctc tttattggct atatgccaat acactgtcct tcagagactg acacggactc 2160  
tgtattttta caggatgggg tctcatttat tatttcaaaa ttcacatata caacaccacc 2220  
gtccccagtg cccgcagttt ttattaaaca taacgtggga tctccacgcg aatctcgggt 2280  
acgtgttccg gacatgggct ctctccgggt agcggcgagg cttctacatc cgagccctgc 2340  
tcccatgcct ccagcgactc atggtcgctc ggcagctcct tgctcctaac agtggaggcc 2400  
agacttaggc acagcacgat gccaccacc accagtgtgc cgcacaaggc cgtggcggtg 2460  
gggtatgtgt ctgaaaaatga gctcggggag cgggcttgca ccgctgacgc atttggaaga 2520  
cttaaggcag cggcagaaga agatgcaggc agctgagttg ttgtgttctg ataagagtca 2580  
gaggtaaact ccgttgcggt gctgttaacg gtggagggca gtgtagtctg agcagtagtc 2640  
gttgctgccg cgcgcgccac cagacataat agctgacaga ctaacagact gttcctttcc 2700  
atgggtcttt tctgcagtc cccgggggat ccttcgaacg tagctctaga ttgagtcgac 2760  
gttactggcc gaagccgctt ggaataaggc cgggtgtcgt ttgtctatat gttattttcc 2820  
accatattgc cgtcttttgg caatgtgagg gcccggaac ctggccctgt cttcttgacg 2880  
agcatctcta ggggtctttc cctctctgcc aaaggaatgc aaggctctgt gaatgtcgtg 2940  
aaggaagcag ttccctctga agcttcttga agacaaacaa cgtctgtagc gaccctttgc 3000  
aggcagcgga accccccacc tggcgacagc tgctctgcg gccaaaagcc acgtgtataa 3060  
gatacacctg caaaggcggc acaaccccag tgccacgttg tgagttggat agttgtggaa 3120  
agagtcaaat ggctctctc aagcgtattc aacaaggggc tgaaggatgc ccagaaggtg 3180  
ccccattgta ttgggtctga tctggggcct gctttacatg gcttttagtc 3240  
aggttaaaaa aacgtctagg cccccgaac cacggggacg tggttttcct ttgaaaaaca 3300  
cgataatacc atgggtaagt gatattctat agttgtgacc ggcgcctagt gttgacaatt 3360  
aatcatcggc atagtatatc aatacgactc actataggag ggcaccatg 3420  
tcgactacta accttcttct ctttctctaca gctgagatca ccggtaggag ggccatcatg 3480

aaaaagcctg	aactcaccgc	gacgtctgtc	gcgaagtttc	tgatcgaaaa	gttcgacagc	3540
gtctccgacc	tgatgcagct	ctcggagggc	gaagaatctc	gtgctttcag	cttcgatgta	3600
ggagggcggtg	gatatgtcct	gcggtgaaat	agctgcgcgc	atggtttcta	caaagatcgt	3660
tatgttttatc	ggcactttgc	atcggccgcg	ctcccgatct	cggaagtgtc	tgacattggg	3720
gaattcagcg	agagcctgac	ctattgcatc	tcccgcctgt	cacagggtgt	cacgttgcaa	3780
gacctgcctg	aaaccgaact	gccccgtggt	ctgcaaccgc	tgcggagctc	catggatgcg	3840
atcgctgcgg	ccgatcttag	ccagacgagc	gggttcggcc	cattcggacc	gcaaggaatc	3900
gggtcaataca	ctacatggcg	tgatttcata	tgcgcgattg	ctgatcccca	tgtgtatcac	3960
tggcaaaactg	tgatggacga	caccgtcagt	gcgtccgtcg	cgcaggctct	cgatgagctg	4020
atgctttggg	ccgaggactg	ccccgaagtc	cggcacctcg	tgacgcggga	tttcggctcc	4080
aacaatgtcc	tgacggacaa	tggccgcata	acagcgggtc	ttgactggag	cgaggcgatg	4140
ttcgggggatt	cccaatacga	ggtcgcccaac	atcttcttct	ggaggccgtg	gttggcttgt	4200
atggagcagc	agacgcgcta	cttcgagcgg	aggcatccgg	agcttgacgg	atcgcccgcc	4260
ctccggggcgt	atatgctccg	cattgggtctt	gaccaactct	atcagagctt	ggttgacggc	4320
aatttcgatg	atgcagcttg	ggcgcagggt	cgatgcgacg	caatcgtccg	atccggagcc	4380
gggactgtcg	ggcgtacaca	aatcgcccg	agaagcgcg	ccgtctggac	cgatggctgt	4440
gtagaagtag	tcgccgatag	tggaaaccga	cgccccagca	ctcgtccgag	ggcaaaggaa	4500
tgagtcgaga	attcgctaga	gggcccattt	ctatagtgtc	acctaaatgc	tagagctcgc	4560
tgatcagcct	cgactgtgcc	ttctagtgtc	cagccatctg	ttgtttgccc	ctcccccggtg	4620
ccttccttga	ccctggaagg	tgccactccc	actgtccttt	cctaataaaa	tgaggaaatt	4680
gcatcgcat	gtctgagtag	gtgtcattct	attctggggg	gtgggggtggg	gcaggacagc	4740
aaggggggagg	attgggaaga	caatagcagg	catgcgcagg	gcccatttgc	tcgagcggcc	4800
gcaataaaat	atctttatatt	tcattacatc	tgtgtgttgg	ttttttgtgt	gaatcgtaac	4860
taacatacgc	tctccatcaa	aacaaaacga	aacaaaacaa	actagcaaaa	taggctgtcc	4920
ccagtgaag	tgcaggtgcc	agaacatttc	tctatcgaag	gatctgcgat	cgctccgggtg	4980
cccgctcagtg	ggcagagcgc	acatcgccca	cagtccccg	gaagttgggg	ggaggggtcg	5040
gcaattgaac	cggtgcctag	agaaggtggc	gcggggtaaa	ctgggaaagt	gatgtcgtgt	5100
actggctccg	cctttttccc	gaggggtggg	gagaaccgta	tataagtgc	gtagtccgc	5160
tgaacgttct	tttttcgcaac	gggtttgccc	ccagaacaca	gctgaagctt	cgaggggctc	5220
gcatctctcc	ttcacgcgcc	cgccgcccta	cctgagggcg	ccatccacgc	cggttgagtc	5280
gcgttctgcc	gcctcccgcg	tgtggtgcct	ctgaactgc	gtccgcgctc	taggttaagt	5340
taaagctcag	gtcagagacc	ggcctttgtc	cggcgctccc	ttggagccta	cctagactca	5400
gccggctctc	cacgctttgc	ctgaccctgc	ttgtctcaat	ctacgtcttt	gtttcgtttt	5460
ctgttctgcg	ccgttacaga	tccaagctgt	gaccggcgcc	tacgtaagt	atatctacta	5520
gatttatcaa	aaagagtgtt	gacttgtgag	cgctcacaat	tgatacttag	attcatcgag	5580
agggacacgt	cgactactaa	ccttctttctc	tttctctacg	ctgagat		5627

<210> 28  
 <211> 553  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pMG plasmid from InvivoGen: EMCV IRES sequence

<400> 28						
aacgttactg	gccgaagccg	cttggaataa	ggccgggtgtg	cgtttgtcta	tatgttatatt	60
tccaccatat	tgccgtcttt	tggcaatgtg	agggcccggga	aacctggccc	tgtcttcttg	120
acgagcattc	ctaggggtct	ttcccctctc	gccaaaggaa	tgcaaggctc	gttgaatgtc	180
gtgaaggaag	cagttcctct	ggaagcttct	tgaagacaaa	caacgtctgt	agcgaccctt	240
tgcaggcagc	ggaaccccc	acctggcgac	aggtgcctct	gcggccaaaa	gccacgtgta	300
taagatacac	ctgcaaaggc	ggcacaaccc	cagtgccacg	ttgtgagttg	gatagttgtg	360
gaaagagtca	aatggctctc	ctcaagcgta	ttcaacaagg	ggctgaagga	tgcccagaag	420
gtaccccat	gtatgggata	tgatctgggg	cctcggtgca	catgctttac	gtgtgtttag	480
tcgaggttaa	aaaacgtcta	ggcccccgga	accacgggga	cgtgggtttc	ctttgaaaaa	540
cacgatgata	ata					553

<210> 29  
 <211> 4692  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pDSred1-N1 plasmid from Clontech

<400> 29						
tagttattaa	tagtaataca	ttacgggggtc	attagttcat	agcccatata	tggagttccg	60

cggtacataa	cttacggtaa	atggccccgc	tggttgaccg	cccaacgacc	ccccccatt	120
gacgtcaata	atgacgtatg	ttcccatagt	aacgcccaata	gggactttcc	attgacgtca	180
atgggtggag	tattttacgg	aaactgcccc	cttggcagta	catcaagtgt	atcatatgcc	240
aagtaogccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgcccagta	300
catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac	360
catggtgatg	cggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	420
atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	480
ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatgggcg	gtaggcggtg	540
acgggtgggag	gtctatataa	gcagagctgg	tttagtgaa	cgtagatcc	gctagcgcta	600
ccggactcag	atctcgagct	caagcttcga	attctgcagt	cgacgggtacc	gcgggcccgg	660
gatccaccgg	tcgccaccat	ggtgcgctcc	tccaagaacg	tcatcaagga	gttcatgcgc	720
ttcaaggtgc	gcattggagg	caccgtgaac	ggccacgagt	tcgagatcga	gggcgagggc	780
gaggcccgcc	cctacgaggg	ccacaacacc	gtgaagctga	aggtgaccac	gggcggcccc	840
ctggcccttc	cctggggacat	cctgtccccc	cagttccagt	acggctccaa	ggtgtacgtg	900
aagcaccocg	ccgacattccc	cgactacaag	aagctgtcct	tccccgaggg	cttcaagtgg	960
gagcgcgtga	tgaacttcga	ggacggcgcc	gtggtgaccg	tgaccacagga	ctcctccctg	1020
caggacggct	gcttcatcta	caagggtga	ttcatcgccg	tgaacttccc	ctccgacggc	1080
cccgtaatgc	agaagaagac	catgggctgg	gagggcctca	ccgagcgctc	gtacccccgc	1140
gacggcggtg	tgaagggcga	gatccacaag	gccctgaagc	tgaaggacgg	cgggcactac	1200
ctgggtggag	tcaagttccat	ctacatggcc	aagaagcccg	tgacgctgcc	cggtactact	1260
tacgtggact	ccaagctgga	catcacctcc	cacaacgagg	actacaccat	cgtggagcag	1320
tacgagcgca	ccgaggggccc	ccaccacctg	ttctgtagc	ggccgcgact	ctagatcata	1380
atcagccata	ccacatttgt	agaggtttta	cttgctttta	aaaacctccc	acacctcccc	1440
ctgaacctga	aacataaaat	gaatgcaatt	gttggttgta	acttgtttat	tgacgcttat	1500
aatgggttaca	aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	tttttcactg	1560
cattctagtt	tggtgtttgt	caaactcatc	aatgtatctt	aaggcgtaaa	ttgtaagcgt	1620
taataattttg	ttaaaattcg	cgttaaaatt	ttgttaaatc	agctcatttt	ttaaccaata	1680
ggccgaaatc	ggcaaaatcc	cttataaatc	aaaagaatag	accgagatag	ggttgagtgt	1740
tggtccagtt	tggaacaaga	gtccactatt	aaagaacgtg	gactccaacg	tcaaagggcg	1800
aaaaaccgtc	tatcagggcg	atggcccact	acgtgaacca	tcacccta	caagtttttt	1860
ggggtcgagg	tgccgtaaag	cactaaatcg	gaaccctaaa	gggagccccc	gatttagagc	1920
ttgacgggga	aagcggcgga	acgtggcgag	aaaggaaggg	aagaaagcga	aaggagcggg	1980
cgctagggcg	ctggcaagtg	tagcgggtcac	gctgcgcgta	accaccacac	ccgcgcgcgt	2040
taatgcgcgg	ctacagggcg	cgtagggtgg	cacttttcgg	ggaaatgtgc	gcggaacccc	2100
tattttgttta	tttttctaaa	tacattcaaa	tatgtatccg	ctcatgagac	aataaccctg	2160
ataaatgctt	caataatatt	gaaaaaggaa	gagtcctgag	gcggaagaa	ccagctgtgg	2220
aatgtgtgtc	agttagggtg	tggaaagtcc	ccaggctccc	cagcaggcag	aagtatgcaa	2280
agcatgcata	tcaattagtc	agcaaccagg	tgtggaaagt	ccccaggctc	ccagcaggcc	2340
agaagtatgc	aaagcatgca	tctcaattag	tcagcaacca	tagtcccggc	cctaactccg	2400
cccatcccg	ccctaactcc	gcccagttcc	gcccattctc	cgccccatgg	ctgactaatt	2460
ttttttat	atgcagggc	cgaggccgcc	ctggcctctg	agctattcca	gaagtatga	2520
ggaggctttt	ttggaggcct	aggcttttgc	aaagatcgat	caagagacag	gatgaggatc	2580
gtttgcgatg	attgaacaag	atggattgca	cgcaggttct	ccggccgctt	gggtggagag	2640
gctattcggc	tatgactggg	cacaacagac	aatcggctgc	tctgatgcgc	ccgtgtccg	2700
gctgtcagcg	cagggggcgcc	cggttctttt	tgtcaagacc	gacctgtccg	gtgccctgaa	2760
tgaactgcaa	gacgaggcag	cgcggtatc	gtggctggcc	acgacgggcg	ttccttgccg	2820
agctgtgctc	gacgttgtca	ctgaagcggg	aagggactgg	ctgctattgg	gcgaagtgcc	2880
ggggcaggat	ctcctgtcat	ctcaccttgc	tcctgccgag	aaagtatcca	tcatggctga	2940
tgcaatgcgg	cggctgcata	cgcttgatcc	ggctacctgc	ccattcgacc	accaagcgaa	3000
acatcgcatc	gagcagcac	gtactcggat	ggaagccggg	cttgtcgatc	aggatgatct	3060
ggacgaagag	catcaggggc	tcgcgccagc	cgaactgttc	gccaggctca	aggcgagcat	3120
gcccgaacgg	gaggatctcg	tcgtgaccac	tggcgatgcc	tgcttgccga	atatcatggt	3180
ggaaaatggc	cgcttttctg	gattcatcga	ctgtggccgg	ctgggtgtgg	cggaccgcta	3240
tcaggacata	gcgttggcta	cccgtgatat	tgctgaagag	cttggcgggc	aatgggctga	3300
cogcttcttc	gtgctttacg	gtatcgccgc	tcocgatcgc	cagcgcatcg	ccttctatcg	3360
ccttcttgac	gagttcttct	gagcgggact	ctgggggtcg	aaatgaccga	ccaagcgacg	3420
cccaacctgc	catcacgaga	tttcgattcc	accggcgcc	tctatgaaag	gttgggcttc	3480
ggaatcgttt	tcggggacgc	cggctggatg	atcctccagc	gcgggggatct	catgctggag	3540
ttcttcgccc	accctaaggg	gaggctaact	gaaacacgga	aggagacaat	accggaagga	3600
accgcgcta	tgacggcaat	aaaaagacag	aataaaacgc	acgggtgttg	gtcgtttgtt	3660
cataaacgcg	gggttcgggtc	ccagggctgg	cactctgtgc	ataccccacc	gagaccccat	3720
tggggccaat	acgcccgcgt	ttcttctctt	tccccacc	accccccaag	ttcgggtgaa	3780
ggcccagggc	tcgcagccaa	cgtcggggcg	gcaggccctg	ccatagcctc	aggttactca	3840
tatatacttt	agattgatgt	aaaacttcat	ttttaattta	aaaggatcta	ggtgaagatc	3900
ctttttgata	atctcatgac	caaaatccct	ttaacgtgag	tttcgttcca	ctgacgctca	3960
gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	tttttctgcg	cgtaactctgc	4020
tgcttgcaaa	caaaaaaacc	accgctacca	gcggtggttt	gtttgcccga	tcaagagcta	4080

ccaactcttt	ttccgaaggt	aactggcttc	agcagagcgc	agataccaaa	tactgtcctt	4140
ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcacccgc	tacatacctc	4200
gctctgctaa	tcctgttacc	agtggctgct	gccagtgccg	ataagtcgtg	tcttaccggg	4260
ttggactcaa	gacgatagtt	accggataag	gcgcagcggg	cgggctgaac	gggggggtcg	4320
tgacacagc	ccagcttggg	gcgaacgacc	tacaccgaac	tgagatacct	acagcgtgag	4380
ctatgagaaa	gcgccacgct	tcccgaaggg	agaaaggcgg	acaggtatcc	ggtaagcggc	4440
agggtcggaa	caggagagcg	cacgagggag	cttccagggg	gaaacgcctg	gtatctttat	4500
agtcctgtcg	ggtttcgcca	cctctgactt	gagcgtcgat	ttttgtgatg	ctcgtcaggg	4560
gggcggagcc	tatggaaaaa	cgccagcaac	gcgccctttt	tacggttcct	ggccttttgc	4620
tggccttttg	ctcacatggt	ctttcctgcg	ttatccctcg	attctgtgga	taaccgtatt	4680
accgccatgc	at					4692

<210> 30  
 <211> 4257  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pPur plasmid from Clontech

<400> 30						
ctgtggaatg	tgtgtcagtt	agggtgtgga	aagtcgccag	gctccccagc	aggcagaagt	60
atgcaaaagca	tgcatctcaa	ttagtcagca	accaggtgtg	gaaagtcccc	aggctcccca	120
gcaggcagaa	gtatgcaaag	catgcatctc	aattagtcag	caaccatagt	cccgccctta	180
actccgcccc	tcccgcctct	aactccgccc	agttccgccc	attctccgcc	ccatggttga	240
ctaatttttt	ttattttatgc	agaggccgag	gccgcctcgg	cctctgagct	attccagaag	300
tagtgaggag	gcttttttgg	aggcctaggc	ttttgcaaaa	agcttgcatg	cctgcaggtc	360
ggccgcccag	accggtgccc	ccaccatccc	ctgacccacg	cccctgaccc	ctcacaagga	420
gacgaccttc	catgaccgag	tacaagcccc	cgggtgcgct	cgccaccgcg	gacgacgtcc	480
cccgggccgt	acgcaccctc	gccgcgcgct	tcgcccacta	ccccgccacg	cgccacaccg	540
tgcaccggga	ccgccacatc	gagcgggtca	ccgagctgca	agaactcttc	ctcacgcgcg	600
tcgggtctga	catcggcaag	gtgtgggtcg	cggacgacgg	cgccgcgggtg	gcggtctgga	660
ccacgcccga	gagcgtcgaa	gcggggggcgg	tggttcgccc	gatcggcccc	cgcatggccg	720
agttgagcgg	ttcccggctg	gccgcgcagc	aacagatgga	aggcctcctg	gcgcgcgacc	780
ggcccaagga	gcccgctggg	ttcctggcca	ccgtcggcgt	ctcgcccgac	caccagggca	840
agggtctggg	cagcgcctgc	gtgctccccg	gagtgagggc	ggccgagcgc	gccgggggtgc	900
ccgccttcct	ggagacctcc	gcgccccgca	acctccccct	ctacgagcgg	ctcgggttca	960
ccgtcaccgc	cgacgtcgag	gtgcccgaag	gaccgcgcac	ctggtgcatg	accgcgaagg	1020
ccgggtgcctg	acgcccgcgc	cacgaccgcg	acgcgccgac	cgaaaaggagc	gcacgacccc	1080
atgggtccga	ccgaagccga	ccggggcggc	cccgccgacc	ccgcacccgc	ccccgaggcc	1140
caccgactct	agaggatcat	aatcagccat	accacatttg	tagaggtttt	acttgcttta	1200
aaaaaacctcc	cacacctccc	cctgaacctg	aaacataaaa	tgaatgcaat	tggttggtgt	1260
aacttggtta	ttgcagctta	taatgggttac	aaataaagca	atagcatcac	aaatttcaca	1320
aataaagcat	ttttttcact	gcattctagt	tgtggtttgt	ccaaactcat	caatgtatct	1380
tatcatgtct	ggatccccag	gaagctcctc	tgtgtcctca	taaaccctaa	cctcctctac	1440
ttgagaggac	attccaatca	taggctgccc	atccaccctc	tgtgtcctcc	tggttaattag	1500
gtcacttaac	aaaaaggaaa	ttgggttaggg	gtttttcaca	gaccgctttc	taagggtaat	1560
tttaaaatat	ctgggaagtc	ccttccactg	ctgtgttcca	gaagtgttgg	taaacagccc	1620
acaaatgtca	acagcagaaa	catacaagct	gtcagctttg	cacaagggcc	caacaccctg	1680
ctcatcaaga	agcactgtgg	ttgctgtggt	agtaatgtgc	aaaacaggag	gcacattttc	1740
cccacctgtg	taggttccaa	aatatctagt	gttttcattt	ttacttggat	caggaaccca	1800
gcactccact	ggataagcat	tatccttatc	caaaacagcc	ttgtggtcag	tggtcatctg	1860
ctgactgtca	actgtagcat	tttttggggg	tacagtttga	gcaggatatt	tggtcctgta	1920
gtttgtctaac	acaccctgca	gctccaaagg	ttccccacca	acagcaaaaa	aatgaaaatt	1980
tgacccttga	atgggttttc	cagcaccatt	ttcatgagtt	ttttgtgtcc	ctgaatgcaa	2040
gtttaacata	gcagttaccc	caataacctc	agttttaaca	gtaacagctt	cccacatcaa	2100
aataatttcca	cagggttaagt	cctcatttaa	attaggcaaa	ggaattcttg	aagacgaaag	2160
ggcctcgtga	tacgcctatt	tttataggtt	aatgtcatga	taataatggg	ttcttagacg	2220
tcagggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	tttgtttatt	tttctaataa	2280
cattcaaaata	tgtatccgct	catgagacaa	taacctgat	aaatgcttca	ataaatgtga	2340
aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgcgc	ttattccctt	ttttgcgcca	2400
ttttgccttc	ctgtttttgc	tcacccagaa	acgctgggtga	aagtaaaaaga	tgctgaagat	2460
cagttgggtg	ccagagtggg	ttacatcgaa	ctggatctca	acagcggtaa	gatccttgag	2520
agttttcgcc	ccgaagaacg	ttttccaatg	atgagcactt	ttaaagtctt	gctatgtggc	2580
gcggtattat	cccggtgttga	cgccggggcaa	gagcaactcg	gtcgccgcat	acactatttc	2640
cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	tggtcatgca	2700
gtaagagaat	tatgcagtgc	tgccataacc	atgagtgtata	acactgcggc	caacttactt	2760



ctgacaacga	tccggaggacc	gaaggagcta	accgcttttt	tgcacaacat	gggggatcat	2820
gtaactcgcc	ttgatcggtg	ggaaccggag	ctgaatgaag	ccataccaaa	cgacgagcgt	2880
gacaccacga	tgccctgcagc	aatggcaaca	acgttgcgca	aactattaac	tggcgaacta	2940
cttactctag	cttcccggca	acaaltaata	gactggatgg	aggcggataa	agttgcagga	3000
ccacttctgc	gctcggccct	tccggctggc	tggtttattg	ctgataaatc	tggagccggg	3060
gagcgtgggt	ctcggcggtat	cattgcagca	ctggggccag	atggtaagcc	ctcccgtatc	3120
gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	acagatcgct	3180
gagatagggt	cctcactgat	taagcattgg	taactgtcag	accaagttta	ctcatatata	3240
cttttagattg	atttaaaact	tcatttttta	tttaaaaagg	tctagggtgaa	gatccttttt	3300
gataatctca	tgacccaaa	cccttaacgt	gagttttcgt	tccactgagc	gtcagacccc	3360
gtagaaaaa	tcaaaggatc	ttcttgagat	cccttttttc	tgccgcta	ctgctgcttg	3420
caaacaaaa	aaccaccgct	accagcgggt	gtttgtttgc	cggatcaaga	gctaccaact	3480
ctttttccga	aggttaactgg	cttcagcaga	gcgcagatac	caaatactgt	ccttctagt	3540
tagccgtagt	taggccacca	cttcaagaac	tctgtagcac	cgccctacata	cctcgctctg	3600
ctaactcctgt	taccagtggc	tgctgccagt	ggcgataagt	cgtgtcttac	cgggttggac	3660
tcaagacgat	agttaccgga	taaggcgag	cggtcgggct	gaacgggggg	ttcgtgcaca	3720
cagcccagct	tggagcgaa	gacctacacc	gaactgagat	acctacagcg	tgagctatga	3780
gaaagcgcca	cgcttcccga	agggagaaag	gcggacaggt	atccggtaag	cggcagggtc	3840
ggaacaggag	agcgacagag	ggagcttcca	gggggaaacg	cctgggtatct	ttatagtctt	3900
gtcgggtttc	gccacctctg	acttgagcgt	cgatttttgt	gatgctcgtc	agggggggcg	3960
agcctatgga	aaaacgccag	caacgcggcc	tttttacggg	tcctggcctt	ttgctggcct	4020
tttgctcaca	tggtctttcc	tgcgttatcc	cctgattctg	tggaataaccg	tattaccgcc	4080
tttgagtgag	ctgataccgc	tcgcccagc	cgaacgacgg	agcgacgca	gtcagtgagc	4140
gaggaagcgg	aagagcgct	gatgcgggat	tttctcctta	cgcctctgtg	cggattttca	4200
caccgcata	ggtgcactct	cagtacaatc	tgctctgatg	ccgcatagtt	aagccag	4257

<210> 31  
 <211> 8136  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pWE15 cosmid vector

<300>  
 <308> GenBank X65279  
 <309> 1995-04-14

<400> 31						
ctatagtgag	tcgtattatg	cggccgcgaa	ttcttgaaga	cgaaaggggc	tcgtgatacg	60
cctattttta	taggttaatg	tcctgataat	aatggtttct	tagacgtcag	gtggcacttt	120
tcggggaaat	gtgcgcggaa	cccctatttg	tttatttttc	taaatacatt	caaataatgta	180
tccgctcatg	agacaataac	cctgataaat	gcttcaataa	tattgaaaaa	ggaagagtat	240
gagtattcaa	catttccgtg	tcgcccttat	tccttttttt	gcggcatttt	gcttccgtgt	300
tttgctcacc	cagaaaacgt	ggtgaaaagta	aaagatgctg	aagatcagtt	gggtgcacga	360
gtgggttaca	tcgaactgga	tctcaacagc	ggtaagatcc	ttgagagttt	tcgccccgaa	420
gaacggtttc	caatgatgag	cactttttaa	gttctgctat	gtggcgcggg	attatccggt	480
gttgacgcgg	ggcaagagca	actcggtcgc	cgcatacact	attctcagaa	tgacttggtt	540
gagtactcac	cagtcacaga	aaagcatctt	acggatggca	tgacagtaag	agaattatgc	600
agtgtgtcca	taaccatgag	tgataaacct	gcggccaact	tacttctgac	aacgatcggg	660
ggaccgaagg	agctaaccgc	ttttttgcac	aacatggggg	atcatgtaac	tcgccttgat	720
cgttggggaa	cggagctgaa	tgaagccata	ccaaacgacg	agcgtgacac	cacgatgcct	780
gcagcaaatg	caacaacgtt	gcgcaaaact	ttaactggcg	aactacttac	tctagcttcc	840
cggcaacaat	taatagactg	gatggaggcg	gataaaagt	caggaccact	tctgcgctcg	900
gcccttccgg	ctggctgggt	tattgctgat	aaatctggag	ccggtgagcg	tgggtctcgc	960
ggtatcattg	cagcactggg	gccagatggg	aagccctccc	gtatcgtagt	tatctacacg	1020
acggggagtc	aggcaactat	ggatgaacga	aatagacaga	tcgctgagat	aggtgcctca	1080
ctgattaagc	atttgtaact	gtcagaccaa	gtttactcat	atatacttta	gattgattta	1140
aaacttcatt	tttaatttaa	aaggatctag	gtgaagatcc	tttttgataa	tctcatgacc	1200
aaaatccctt	aacgttgagtt	ttcgttccac	tgagcgtcag	accccgtaga	aaagatcaaa	1260
ggatcttctt	gagatccttt	ttttctgcgc	gtaatctgct	gcttgcaaac	aaaaaaacca	1320
ccgctaccag	cgggtgggtt	tttgccggat	caagagctac	caactccttt	tccgaaggta	1380
actggcttca	gcagagcgca	gataccaaat	actgtccttc	tagtgtagcc	gtagttaggc	1440
caccacttca	agaactctgt	agcaccgcct	acatacctcg	ctctgcta	cctgttacca	1500
gtggctgtcg	ccagtggcga	taagtcgtgt	cttaccgggt	tggaactcaag	acgatagtta	1560
ccgataaagg	cgacgcggtc	gggctgaacg	gggggttcgt	gcacacagcc	cagcttggag	1620
cgaacgacct	acaccgaact	gagataccta	cagcgtgagc	tatgagaaag	cgccacgctt	1680



ccgaaggggag	aaagggcggac	aggatatccgg	taagcggcgag	ggtcgggaaca	ggagagcgca	1740
cgaggggagct	tccagggggga	aacgcctggg	atcttttatag	tcctgtcggg	gtttcgccac	1800
ctctgacttg	agcgtcgatt	tttgtgatgc	tcgtcagggg	ggcggagcct	atggaaaaac	1860
gccagcaacg	cggccttttt	acggttcctg	gccttttgct	ggccttttgc	tcacatgttc	1920
tttctgctg	tatccccctga	ttctgtggat	aaccgtatta	ccgcctttga	gtgagctgat	1980
accgctcgcc	gcagccgaac	gaccgagcgc	agcaggtcag	tgagcgagga	agcgggaagag	2040
cgctgacttc	cgcgtttcca	gactttacga	aacacggaaa	ccgaagacca	ttcatgttgt	2100
tgctcagggt	gcagacgttt	tgacgacgca	gtcgtttcac	gttcgctcgc	gtatcgggtga	2160
ttcattctgc	taaccagtaa	ggcaaccccc	ccagcctagc	cgggtcctca	acgacaggag	2220
cacgatcatg	cgcacccgct	agatccagac	atgataagat	acattgatga	gtttggacaa	2280
accacaacta	gaatgcagtg	aaaaaaatgc	tttattttgtg	aaatttgtga	tgctattgct	2340
ttattttgtaa	ccattataag	ctgcaataaaa	caagttaaca	acaacaattg	cattcatttt	2400
atgttttcagg	ttcagggggga	gggtgtgggag	gttttttaaaa	gcaagtataaa	cctctacaaa	2460
tggtggtatgg	ctgattatga	tctctagtca	aggcactata	catcaaatat	tccttattaa	2520
ccccttttaca	aattaaaaaag	ctaaaggtac	acaatttttg	agcatagtta	ttaatagcag	2580
acactctatg	cctgtgtgga	gtaagaaaaa	acagtatgtt	atgattataa	ctgttatgcc	2640
tacttataaaa	ggttacagaa	tattttttcca	taatttttctt	gtatagcagt	gcagcttttt	2700
cctttgtggt	gtaaatagca	aagcaagcaa	gagttctatt	actaaacaca	gcatgactca	2760
aaaaacttag	caattctgaa	ggaaagtcct	tgggttcttc	taccttttctc	ttcttttttg	2820
gaggagttag	atgttgagag	tcagcagtag	cctcatcatc	actagatggc	atttcttctg	2880
agcaaaacag	gttttcctca	ttaaaggcat	tccaccactg	ctcccattca	tcagttccat	2940
agggttggat	ctaaaatata	caaacaatta	gaatcagtag	tttaacacat	tatacattta	3000
aaaattttat	atttacctta	gagctttaaa	tctctgtagg	tagtttgtcc	aattatgtca	3060
caccacagaa	gtaagggtcc	ttcacaaaaga	tccggaccac	agcggccatc	gtgctcccc	3120
actcctgcag	ttcggggggca	tggtatgcgcg	gatagccgct	gctgggttcc	tggtatgcga	3180
cggatttgca	ctgcgcgtag	aactcgcgag	gtcgtccagc	ctcaggcagc	agctgaacca	3240
actcgcgagg	ggatcgagcc	cgggggtgggc	gaagaactcc	agcatgagat	ccccgcgctg	3300
gaggatcatc	cagccggcgt	cccggaaaaac	gattccgaag	cccaaccttt	catagaaggc	3360
ggcgggtgga	tcgaaatctc	gtgatggcag	gttgggcgtc	gcttgggtcgg	tcatttcgaa	3420
cccagagtc	ccgctcagaa	gaactcgtca	agaaggcgat	agaaggcgat	gcgctgcgaa	3480
tcgggagcgg	cgataccgta	aagcacgagg	aagcgggtcag	cccattcgcc	gccaagctct	3540
tcaggaatat	caacggtagc	caacgctatg	tccgtatagc	ggtcgcgcac	accagccgg	3600
ccacagtcga	tgaatccaga	aaagcggcca	ttttccacca	tgatattcgg	caagcaggca	3660
tcgcctatgg	tcacgacgag	atcctcgccg	tccgggatgc	cgccttgagc	ctggcgaaca	3720
gttcgggtgg	cgcgcggtag	tgatgctctt	cgctccagac	atcctgatcg	acaagaccgg	3780
cttccatccg	agtagctgct	cgctcgatgc	gatgtttcgc	ttgggtggtcg	aatgggcagg	3840
tagccggatc	aagcgtatgc	agccgcgcga	ttgcattcag	catgatggat	actttctcgg	3900
caggagcaag	gtgcagtgac	aggagatcct	gtcccgccac	ttcgcccaat	agcagccagt	3960
cccttcccgc	cgcgcgtgcc	acgtcgagca	cagctgcgca	aggaacgcgc	gtcgtggcca	4020
gccacgatag	ccgcgcgtgc	tcgtcctgca	gttcattcag	ggcaccggac	aggtcggtct	4080
tgacaaaaag	acccggcgcc	ccctgcgctg	acagccggaa	cacggcggca	tcagagcagc	4140
cgatattgctg	ttgtgcccag	tcatagccga	atagcctctc	caccacaaggc	gcccggagaac	4200
ctgcgtgcaa	tccattctgt	tcaatcatgc	gaaacgatcc	tcattcctgtc	tcttgatcag	4260
atcttgatcc	ccctgcgccat	cagatccttg	gcggcaagaa	agccatccag	ttacttttgc	4320
agggtttccc	aaccttacc	gagggcgccc	cagctggcaa	ttccgggttcg	cttgcgtgctc	4380
ataaaaccgc	ccagttatgc	tatcgccatg	taagccctac	gcaagctacc	tgctttctct	4440
ttgcgcttgc	gttttccctt	gtccagatag	cccagtagct	gacattcctc	cggggtcagc	4500
accgtttctg	cggaactggct	ttctacgtgt	tccgcttccct	ttagcagccc	ttgcgcctcg	4560
agtgccttgcg	gcagcgtgaa	agctttttgc	aaaagcctag	gcctccaaaa	aagcctcctc	4620
actacttctg	gaatagctca	gagggcggag	cggcctaaat	aaaaaaaatt	agtcagccat	4680
ggggcgggaga	atgggcggga	ctgggcggag	ttaggggcgg	gatgggcgga	gttaggggag	4740
ggactatggt	tgctgactaa	ttgagatgca	tgctttgcat	acttctgcct	gctggggagc	4800
ctggggactt	tccacacctg	gttgctgact	aattgagatg	catgctttgc	atacttctgc	4860
ctgctgggga	gcctggggac	tttccacacc	ctaactgaca	cacattccac	agccggatct	4920
gcaggaccca	acgctgcccg	agatgcgcgc	cgtgcggctg	ctggagatgg	cggacgcgat	4980
ggatatgttc	tgccaagggt	tggtttgcgc	attcacagtt	ctccgcaaga	attgattggc	5040
tccaattctt	ggagtgggtga	atccgttagc	gaggtgcgcg	cggcttccat	tcaggtcgag	5100
gtggcccggc	tccatgcacc	gcgacgcacc	gcgggggagc	agacaaggta	tagggcgggc	5160
cctacaatcc	atgccaaccc	gttccatggt	ctgcgcggag	cgcataaatc	gccgtgacga	5220
tcagcgggtcc	aatgatcgaa	gttaggctgg	taagagccgc	gagcgatcct	tgaagctgtc	5280
cctgatggtc	gtcatctacc	tgccctggaca	gcattggcctg	caacgcggca	tcccgatgcc	5340
gccggaagcg	agaagaatca	taattggggaa	ggccatccag	cctcgcgtcg	cgaacgccag	5400
caagacgtag	cccagcgcgt	cgggcgcgca	tgccgcgcag	aatggcctgc	ttctcgccga	5460
aacgtttggt	ggcggggacca	gtgacgaagg	cctgagcgag	ggcgtgcaag	attccgaata	5520
ccgcaagcga	caggccgatc	atcgtcgcgc	tcagacgaaa	gcggtcctcg	ccgaaaaatga	5580
ccagagcgc	tgccggcacc	tgctcctacga	gttgcatgat	aaagaagaca	gtcataagtg	5640
cggcgacgat	agtcatgccc	cgcgccccacc	ggaaggagct	gactgggttg	aaggctctca	5700

agggcatcgg	tgcagcgtct	cccttatgcg	actcctgcat	taggaagcag	cccagtagta	5760
ggttgaggcc	gttgagcacc	gccgcgcgca	ggaatgggtgc	atgcaaggag	atggcgccca	5820
acagtccccc	ggccacgggg	ctgccaccat	acccacgccc	aaacaaggcg	tcagtagccc	5880
gaagtggcga	gcccgatctt	ccccatcggt	gatgtcggcg	atataggcgc	cagcaaccgc	5940
acctgtggcg	ccggtgatgc	cggccacgat	gcgtccggcg	tagaggatct	tggcagtcac	6000
agcatgcgca	tatccatgct	tcgaccatgc	gctcacaag	taggtgaatg	cgcaatgtag	6060
taccacatc	gtcatcgctt	tccactgctc	tcgcgaataa	agatggaaaa	tcaatctcat	6120
ggtaatagtc	aactaaaatc	cttgtattca	taaatcctcc	aggtagctat	atgcaaattg	6180
aaacaaaaga	gatgggtgatc	tttctaagag	atgatggaa	ctcccttcag	tatcccgatg	6240
gtcaatgcgc	tggatatggg	atagatggga	atatgctgat	ttttatggga	cagagttgcg	6300
aactgttccc	aactaaaatc	atgttgacag	atcagctcac	tacgaacttt	acccacaaat	6360
agtcaggtaa	tgaatcctga	tataaagaca	ggttgataaa	tcagtcttct	acgcgcacgc	6420
cacgcgcaca	ccgtagaaaag	tcttttcagtt	gtgagcctgg	gcaaaccggt	aactttcggc	6480
ggctttgctg	tgcgacaggc	tcacgtctaa	aaggaaaataa	atcatgggtc	ataaaattat	6540
cacgttgtcc	ggcgcggcga	cggatgttct	gtatgcgctg	tttttccgtg	gcgcgttgct	6600
gtctgggtgat	ctgccttcta	aatctggcac	agccgaattg	cgcgagcttg	gttttgtgta	6660
aaccagacac	acagaaaatc	aataccagaa	agaaaaatcac	tttacctttc	tgacatcaga	6720
agggcagaaa	tttgccgttg	aacacctggt	caatacgcgt	tttggtgagc	agcaatattg	6780
cgcttcgatg	acgcttgagg	ttgagattga	tacctctgct	gcacaaaagg	caatcgacga	6840
gctggaccag	cgcattcggt	acaccgtctc	cttcgaactt	attcgcaatg	gagtgctcatt	6900
catcaaggac	gccgctatcg	caaattggtgc	tatccacgca	gcggcaatcg	aaacacctca	6960
gccggtgacc	aatatctaca	acatcagcct	tggtatccag	cgtgatgagc	cagcgcagaa	7020
caaggtaacc	gtcagtgccg	ataagttcaa	agttaaacct	ggtgttgata	ccaacattga	7080
aacgttgatc	gaaaacgcgc	tgaaaaacgc	tgctgaatgt	gcggcgctgg	atgtcacaaa	7140
gcacaaatggc	gcagacaaga	aagcgatgga	tgaactgggt	tcctatgtcc	gcacggccat	7200
catgatggaa	tggttccccg	gtggtgttat	ctggcagcag	tgccgtcgat	agtatgcaat	7260
tgataattat	tatcatttgc	gggtcctttc	cggcgatccg	ccttgttacg	ggcgcgcgac	7320
ctcgcgggtt	ttcgctatatt	atgaaaaattt	tccgggttaa	ggcgtttccg	ttcttcttcg	7380
tcataactta	atgtttttat	ttaaaaatacc	ctctgaaaag	aaaggaaaacg	acaggtgctt	7440
aaagcgagct	ttttggcctc	tgctgtttcc	tttctctgtt	tttgtccgtg	gaatgaacaa	7500
tggaagtcaa	caaaaagcag	ctgggtgaca	ttttcgggtgc	gagtatccgt	accattcaga	7560
actggcagga	cccggtatgc	aaatgggtatg	ccgaaaaggga	tgctgaaatt	gagaacgaaa	7620
atgactctgc	cgccgtcata	ggaggttgaa	aggccagcga	ggcagatcca	caggacgggt	7680
agctgcgcgc	tgatcgcgta	gtcgatagtg	gctccaagta	gcgaagcag	caggatggg	7740
gtggtcgcca	gcggctcgac	agtgctccga	gaacgggtgc	gcatagaaat	tgcatcaacg	7800
cgccggcaca	tagcagcacg	ccatagtgag	tgccgatgct	gtcgggaatgg	acgatatccc	7860
catatagcgc	cggcagtagc	ggcataacca	agcttatgcc	tacagcatcc	agggtagcgg	7920
gcaagaggcc	cgccagtagc	gcattgttag	atctcataca	cggtgcctga	ctgcgttagc	7980
tgccgaggat	tgataaacta	ccgcattaaa	gcttatcgat	gataagcggg	caaacatgag	8040
aatttaactg	cgcaattaac	cctcactaaa	ggatcc			8100
aattcgcgcc						8136

<210> 32  
 <211> 2713  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pNEB193 plasmid

<400> 32						60
tgcgcggttt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacgggtca	120
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcaggggcgcg	tcagcgggtg	180
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	240
accatattgcg	gtgtgaaata	ccgcacagat	gcgtaaggag	aaaataccgc	atcaggcgcc	300
attcgccatt	caggctgcgc	aactgttggg	aagggcgac	ggtgcggggc	tcttcgctat	360
tacgcagcgt	ggcgaagggg	ggatgtgctg	caaggcgatt	aagttgggta	acgccagggt	420
tttcccagtc	acgacgttgt	aaaacgacgg	ccagtgaatt	cgagctcggt	acccgggggc	480
gcgcgggatc	cttaattaa	tctagagtgc	actgtttaaa	cctgcaggca	tgcaagcttg	540
gcgtaatcat	ggtcatagct	gtttcctgtg	tgaaattggt	atccgctcac	aattccacac	600
aacatacgag	ccggaagcat	aaagtgtaaa	gcctgggggtg	cctaatagagt	gagctaactc	660
acattaattg	cgttgcgctc	actgcccgtc	ttccagtcgg	gaaacctgtc	gtgccagctg	720
cattaatgaa	tcggccaacg	cgccggggaga	ggcggtttgc	gtattggggc	ctcttcgct	780
tcctcgctca	ctgactcgct	gcgctcggtc	gttcggctgc	ggcgagcggg	atcagctcac	840
tcaaaaggcgg	taatacgggt	atccacagaa	tcaggggata	acgcaggaaa	gaacatgtga	900
gcaaaaggcc	agcaaaaggc	caggaaccgt	cggttgctggc	cggttgctggc	gttttccat	960
aggctccgcc	cccctgacga	gcatacacia	aatcgacgct	caagtcagag	gtggcgaaac	

ccgacaggac	tataaagata	ccaggcggtt	ccccctggaa	gctccctcgt	gcgctctcct	1020
gttccgaccc	tgccgcttac	cggatacctg	tccgcctttc	tcccttcggg	aagcgtggcg	1080
ctttctcata	gctcacgctg	taggtatctc	agttcggtgt	aggctcgttcg	ctccaagctg	1140
ggctgtgtgc	acgaaccccc	cgttcagccc	gacogctgcg	ccttatccgg	taactatcgt	1200
cttgagtcca	acccggtaag	acacgactta	tcgccactgg	cagcagccac	tggtaacagg	1260
attagcagag	cgaggatatgt	aggcggtgct	acagagttct	tgaagtgggtg	gcctaactac	1320
ggctacacta	gaaggacagt	atcttggtatc	tgcgctctgc	tgaagccagt	taccttcgga	1380
aaaagagttg	gtagctcttg	atccggcaaaa	caaaccaccg	ctggtagcgg	tggttttttt	1440
gtttgcaagc	agcagattac	gcgcagaaaa	aaaggatctc	aagaagatcc	tttgatcttt	1500
tctacggggg	ctgacgctca	gtggaacgaa	aactcacgtt	aagggatttt	ggatcatgaga	1560
ttatcaaaaa	ggatcttcac	ctagatcctt	ttaaattaaa	aatgaagttt	taaatacaatc	1620
taaagtatat	atgagtaaac	ttgggtctgac	agttaccaat	gcttaatcag	tgaggcacct	1680
atctcagcga	tctgtctatt	tcgttcatcc	atagttgcct	gactccccgt	cgtgtagata	1740
actacgatac	gggagggtct	accatctggc	cccagtgctg	caatgatacc	gcgagaccca	1800
cgctcaccgg	ctccagattt	atcagcaata	aaccagccag	ccggaagggc	cgagcgcaga	1860
agtggctcctg	caactttatc	cgcctccatc	cagtcctatta	attgttgccg	ggaagctaga	1920
gtaagttagtt	cgccagttaa	tagtttgccg	aacgttggtg	ccattgctac	aggcatcgtg	1980
gtgtcacgct	cgctgcttgg	tatggcttca	ttcagctccg	gttcccaacg	atcaaggcga	2040
gttacatgat	cccccatggt	gtgcaaaaaa	gcggttagct	ccttcggtcc	tccgatcgtt	2100
gtcagaagta	agttggccgc	agtggttatca	ctcatgggtta	tggcagcact	gcataattct	2160
cttactgtca	tgccatccgt	aagatgcttt	tctgtgactg	gtgagtactc	aaccaagtca	2220
ttctgagaat	agtgtatgcg	gcgaccgagt	tgctcttgcc	cggcgtcaat	acgggataat	2280
accgcgccac	atagcagaac	tttaaaaagt	ctcatcattg	gaaaacgttc	ttcggggcga	2340
aaactctcaa	ggatcttacc	gctgttgaga	tccagttcga	tgtaaccac	tcgtgcaccc	2400
aactgatctt	cagcatcttt	tactttcacc	agcgtttctg	ggtgagcaaa	aacaggaagg	2460
caaaatgccg	caaaaaagg	aataaggcg	acacggaaat	gttgaatact	catactcttc	2520
ctttttcaat	attattgaag	catttatcag	ggttattgtc	tcatgagcgg	atacatattt	2580
gaatgtat	agaaaaataa	acaaatagg	gttccgcgca	catttccccg	aaaagtgcc	2640
cctgacgtct	aagaaacat	tattatcatg	acattaacct	ataaaaaatag	gcgtatcacg	2700
aggccctttc	gtc					2713

<210> 33  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> attP

<400> 33  
 cagctttttt atactaagtt g 21

<210> 34  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> attB

<400> 34  
 ctgctttttt atactaactt g 21

<210> 35  
 <211> 21  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> attL

<400> 35  
 ctgctttttt atactaagtt g 21

<210> 36  
 <211> 21  
 <212> DNA

<213> Artificial Sequence

<220>

<223> attR

<400> 36

cagctttttt atactaactt g

21

<210> 37

<211> 1071

<212> DNA

<213> Artificial Sequence

<220>

<223> Integrase E174R

<221> CDS

<222> (1)...(1071)

<223> Nucleotide sequence encoding Integrase E147R

<400> 37

atg gga aga agg cga agt cat gag cgc cgg gat tta ccc cct aac ctt	48
Met Gly Arg Arg Ser His Glu Arg Asp Leu Pro Pro Asn Leu	
1 5 10 15	
tat ata aga aac aat gga tat tac tgc tac agg gac cca agg acg ggt	96
Tyr Ile Arg Asn Asn Gly Tyr Tyr Cys Tyr Arg Asp Pro Arg Thr Gly	
20 25 30	
aaa gag ttt gga tta ggc aga gac agg cga atc gca atc act gaa gct	144
Lys Glu Phe Gly Leu Gly Arg Asp Arg Arg Ile Ala Ile Thr Glu Ala	
35 40 45	
ata cag gcc aac att gag tta ttt tca gga cac aaa cac aag cct ctg	192
Ile Gln Ala Asn Ile Glu Leu Phe Ser Gly His Lys His Lys Pro Leu	
50 55 60	
aca gcg aga atc aac agt gat aat tcc gtt acg tta cat tca tgg ctt	240
Thr Ala Arg Ile Asn Ser Asp Asn Ser Val Thr Leu His Ser Trp Leu	
65 70 75 80	
gat cgc tac gaa aaa atc ctg gcc agc aga gga atc aag cag aag aca	288
Asp Arg Tyr Glu Lys Ile Leu Ala Ser Arg Gly Ile Lys Gln Lys Thr	
85 90 95	
ctc ata aat tac atg agc aaa att aaa gca ata agg agg ggt ctg cct	336
Leu Ile Asn Tyr Met Ser Lys Ile Lys Ala Ile Arg Arg Gly Leu Pro	
100 105 110	
gat gct cca ctt gaa gac atc acc aca aaa gaa att gcg gca atg ctc	384
Asp Ala Pro Leu Glu Asp Ile Thr Thr Lys Glu Ile Ala Ala Met Leu	
115 120 125	
aat gga tac ata gac gag ggc aag gcg gcg tca gcc aag tta atc aga	432
Asn Gly Tyr Ile Asp Glu Gly Lys Ala Ala Ser Ala Lys Leu Ile Arg	
130 135 140	
tca aca ctg agc gat gca ttc cga gag gca ata gct gaa ggc cat ata	480
Ser Thr Leu Ser Asp Ala Phe Arg Glu Ala Ile Ala Glu Gly His Ile	
145 150 155 160	
aca aca aac cat gtc gct gcc act cgc gca gca aaa tct aga gta agg	528
Thr Thr Asn His Val Ala Ala Thr Arg Ala Ala Lys Ser Arg Val Arg	
165 170 175	
aga tca aga ctt acg gct gac gaa tac ctg aaa att tat caa gca gca	576
Arg Ser Arg Leu Thr Ala Asp Glu Tyr Leu Lys Ile Tyr Gln Ala Ala	

180										185										190										
gaa	tca	tca	cca	tgt	tgg	ctc	aga	ctt	gca	atg	gaa	ctg	gct	gtt	gtt	624														
Glu	Ser	Ser	Pro	Cys	Trp	Leu	Arg	Leu	Ala	Met	Glu	Leu	Ala	Val	Val															
		195					200					205																		
acc	ggg	caa	cga	gtt	ggt	gat	tta	tgc	gaa	atg	aag	tgg	tct	gat	atc	672														
Thr	Gly	Gln	Arg	Val	Gly	Asp	Leu	Cys	Glu	Met	Lys	Trp	Ser	Asp	Ile															
	210					215					220																			
gta	gat	gga	tat	ctt	tat	gtc	gag	caa	agc	aaa	aca	ggc	gta	aaa	att	720														
Val	Asp	Gly	Tyr	Leu	Tyr	Val	Glu	Gln	Ser	Lys	Thr	Gly	Val	Lys	Ile															
225					230					235					240															
gcc	atc	cca	aca	gca	ttg	cat	att	gat	gct	ctc	gga	ata	tca	atg	aag	768														
Ala	Ile	Pro	Thr	Ala	Leu	His	Ile	Asp	Ala	Leu	Gly	Ile	Ser	Met	Lys															
				245					250					255																
gaa	aca	ctt	gat	aaa	tgc	aaa	gag	att	ctt	ggc	gga	gaa	acc	ata	att	816														
Glu	Thr	Leu	Asp	Lys	Cys	Lys	Glu	Ile	Leu	Gly	Gly	Glu	Thr	Ile	Ile															
			260					265					270																	
gca	tct	act	cgt	cgc	gaa	ccg	ctt	tca	tcc	ggc	aca	gta	tca	agg	tat	864														
Ala	Ser	Thr	Arg	Arg	Glu	Pro	Leu	Ser	Ser	Gly	Thr	Val	Ser	Arg	Tyr															
		275					280					285																		
ttt	atg	cgc	gca	cga	aaa	gca	tca	ggt	ctt	tcc	ttc	gaa	ggg	gat	ccg	912														
Phe	Met	Arg	Ala	Arg	Lys	Ala	Ser	Gly	Leu	Ser	Phe	Glu	Gly	Asp	Pro															
	290					295					300																			
cct	acc	ttt	cac	gag	ttg	cgc	agt	ttg	tct	gca	aga	ctc	tat	gag	aag	960														
Pro	Thr	Phe	His	Glu	Leu	Arg	Ser	Leu	Ser	Ala	Arg	Leu	Tyr	Glu	Lys															
305					310					315					320															
cag	ata	agc	gat	aag	ttt	gct	caa	cat	ctt	ctc	ggg	cat	aag	tcg	gac	1008														
Gln	Ile	Ser	Asp	Lys	Phe	Ala	Gln	His	Leu	Leu	Gly	His	Lys	Ser	Asp															
				325					330					335																
acc	atg	gca	tca	cag	tat	cgt	gat	gac	aga	ggc	agg	gag	tgg	gac	aaa	1056														
Thr	Met	Ala	Ser	Gln	Tyr	Arg	Asp	Asp	Arg	Gly	Arg	Glu	Trp	Asp	Lys															
			340					345					350																	
att	gaa	atc	aaa	taa												1071														
Ile	Glu	Ile	Lys	*																										
			355																											

<210> 38  
 <211> 356  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Integrase E147R

<400> 38  
 Met Gly Arg Arg Arg Ser His Glu Arg Arg Asp Leu Pro Pro Asn Leu  
 1 5 10 15  
 Tyr Ile Arg Asn Asn Gly Tyr Tyr Cys Tyr Arg Asp Pro Arg Thr Gly  
 20 25 30  
 Lys Glu Phe Gly Leu Gly Arg Asp Arg Arg Ile Ala Ile Thr Glu Ala  
 35 40 45  
 Ile Gln Ala Asn Ile Glu Leu Phe Ser Gly His Lys His Lys Pro Leu  
 50 55 60  
 Thr Ala Arg Ile Asn Ser Asp Asn Ser Val Thr Leu His Ser Trp Leu  
 65 70 75 80  
 Asp Arg Tyr Glu Lys Ile Leu Ala Ser Arg Gly Ile Lys Gln Lys Thr

Leu	Ile	Asn	Tyr	Met	Ser	Lys	Ile	Lys	Ala	Ile	Arg	Arg	Gly	Leu	Pro
		100						105					110		
Asp	Ala	Pro	Leu	Glu	Asp	Ile	Thr	Thr	Lys	Glu	Ile	Ala	Ala	Met	Leu
		115					120					125			
Asn	Gly	Tyr	Ile	Asp	Glu	Gly	Lys	Ala	Ala	Ser	Ala	Lys	Leu	Ile	Arg
	130					135					140				
Ser	Thr	Leu	Ser	Asp	Ala	Phe	Arg	Glu	Ala	Ile	Ala	Glu	Gly	His	Ile
	145				150					155				160	
Thr	Thr	Asn	His	Val	Ala	Ala	Thr	Arg	Ala	Ala	Lys	Ser	Arg	Val	Arg
			165						170					175	
Arg	Ser	Arg	Leu	Thr	Ala	Asp	Glu	Tyr	Leu	Lys	Ile	Tyr	Gln	Ala	Ala
		180						185					190		
Glu	Ser	Ser	Pro	Cys	Trp	Leu	Arg	Leu	Ala	Met	Glu	Leu	Ala	Val	Val
	195						200					205			
Thr	Gly	Gln	Arg	Val	Gly	Asp	Leu	Cys	Glu	Met	Lys	Trp	Ser	Asp	Ile
	210					215					220				
Val	Asp	Gly	Tyr	Leu	Tyr	Val	Glu	Gln	Ser	Lys	Thr	Gly	Val	Lys	Ile
	225				230					235				240	
Ala	Ile	Pro	Thr	Ala	Leu	His	Ile	Asp	Ala	Leu	Gly	Ile	Ser	Met	Lys
				245					250					255	
Glu	Thr	Leu	Asp	Lys	Cys	Lys	Glu	Ile	Leu	Gly	Gly	Glu	Thr	Ile	Ile
		260					265						270		
Ala	Ser	Thr	Arg	Arg	Glu	Pro	Leu	Ser	Ser	Gly	Thr	Val	Ser	Arg	Tyr
		275					280					285			
Phe	Met	Arg	Ala	Arg	Lys	Ala	Ser	Gly	Leu	Ser	Phe	Glu	Gly	Asp	Pro
	290					295					300				
Pro	Thr	Phe	His	Glu	Leu	Arg	Ser	Leu	Ser	Ala	Arg	Leu	Tyr	Glu	Lys
	305				310					315				320	
Gln	Ile	Ser	Asp	Lys	Phe	Ala	Gln	His	Leu	Leu	Gly	His	Lys	Ser	Asp
			325						330					335	
Thr	Met	Ala	Ser	Gln	Tyr	Arg	Asp	Asp	Arg	Gly	Arg	Glu	Trp	Asp	Lys
		340					345						350		
Ile	Glu	Ile	Lys												
		355													

<210> 39  
 <211> 876  
 <212> DNA  
 <213> Discosoma species

<220>  
 <221> CDS  
 <222> (45)...(737)  
 <223> Nucleotide sequence encoding red fluorescent protein (FP593)

<300>  
 <308> GenBank AF272711  
 <309> 2000-09-26

<400> 39																			
agtttcagcc	agtgacaggg	tgagctgcc	aggtattctaa	caag	atg	agt	tgt	tcc											56
					Met	Ser	Cys	Ser											
					1														
aag	aat	gtg	atc	aag	gag	ttc	atg	agg	ttc	aag	gtt	cgt	atg	gaa	gga				104
Lys	Asn	Val	Ile	Lys	Glu	Phe	Met	Arg	Phe	Lys	Val	Arg	Met	Glu	Gly				
5					10					15					20				
acg	gtc	aat	ggg	cac	gag	ttt	gaa	ata	aaa	ggc	gaa	ggg	gaa	ggg	agg				152
Thr	Val	Asn	Gly	His	Glu	Phe	Glu	Ile	Lys	Gly	Glu	Gly	Glu	Gly	Arg				
				25					30					35					
cct	tac	gaa	ggg	cac	tgt	tcc	gta	aag	ctt	atg	gta	acc	aag	ggg	gga				200
Pro	Tyr	Glu	Gly	His	Cys	Ser	Val	Lys	Leu	Met	Val	Thr	Lys	Gly	Gly				

40										45										50										
cct	ttg	cca	ttt	gct	ttt	gat	att	ttg	tca	cca	caa	ttt	cag	tat	gga	248														
Pro	Leu	Pro	Phe	Ala	Phe	Asp	Ile	Leu	Ser	Pro	Gln	Phe	Gln	Tyr	Gly															
		55					60					65																		
agc	aag	gta	tat	gtc	aaa	cac	cct	gcc	gac	ata	cca	gac	tat	aaa	aag	296														
Ser	Lys	Val	Tyr	Val	Lys	His	Pro	Ala	Asp	Ile	Pro	Asp	Tyr	Lys	Lys															
	70					75					80																			
ctg	tca	ttt	cct	gag	gga	ttt	aaa	tg	gaa	agg	gtc	atg	aac	ttt	gaa	344														
Leu	Ser	Phe	Pro	Glu	Gly	Phe	Lys	Trp	Glu	Arg	Val	Met	Asn	Phe	Glu															
	85				90				95						100															
gac	gg	ggc	gtg	gtt	act	gta	tcc	caa	gat	tcc	agt	ttg	aaa	gac	ggc	392														
Asp	Gly	Gly	Val	Val	Thr	Val	Ser	Gln	Asp	Ser	Ser	Leu	Lys	Asp	Gly															
				105					110					115																
tgt	ttc	atc	tac	gag	gtc	aag	ttc	att	ggg	gtg	aac	ttt	cct	tct	gat	440														
Cys	Phe	Ile	Tyr	Glu	Val	Lys	Phe	Ile	Gly	Val	Asn	Phe	Pro	Ser	Asp															
			120					125					130																	
gga	cct	gtt	atg	cag	agg	agg	aca	cgg	ggc	tg	gaa	gcc	agc	tct	gag	488														
Gly	Pro	Val	Met	Gln	Arg	Arg	Thr	Arg	Gly	Trp	Glu	Ala	Ser	Ser	Glu															
		135					140					145																		
cgt	ttg	tat	cct	cgt	gat	ggg	gtg	ctg	aaa	gga	gac	atc	cat	atg	gct	536														
Arg	Leu	Tyr	Pro	Arg	Asp	Gly	Val	Leu	Lys	Gly	Asp	Ile	His	Met	Ala															
	150					155					160																			
ctg	agg	ctg	gaa	gga	gyc	ggc	cat	tac	ctc	gtt	gaa	ttc	aaa	agt	att	584														
Leu	Arg	Leu	Glu	Gly	Gly	Gly	His	Tyr	Leu	Val	Glu	Phe	Lys	Ser	Ile															
	165				170					175					180															
tac	atg	gta	aag	aag	cct	tca	gtg	cag	ttg	cca	ggc	tac	tat	tat	gtt	632														
Tyr	Met	Val	Lys	Lys	Pro	Ser	Val	Gln	Leu	Pro	Gly	Tyr	Tyr	Tyr	Val															
				185					190					195																
gac	tcc	aaa	ctg	gat	atg	acg	agc	cac	aac	gaa	gat	tac	aca	gtc	gtt	680														
Asp	Ser	Lys	Leu	Asp	Met	Thr	Ser	His	Asn	Glu	Asp	Tyr	Thr	Val	Val															
			200					205					210																	
gag	cag	tat	gaa	aaa	acc	cag	gga	cgc	cac	cat	ccg	ttc	att	aag	cct	728														
Glu	Gln	Tyr	Glu	Lys	Thr	Gln	Gly	Arg	His	His	Pro	Phe	Ile	Lys	Pro															
		215					220					225																		
ctg	cag	tga	actcggctca	gtcatggatt	agcggtaatg	gccacaaaag										777														
Leu	Gln	*																												
	230																													
gcacgatgat	cgtttttttag	gaatgcagcc	aaaaattgaa	ggttatgaca	gtagaaatac	837																								
aagcaacagg	ctttgcttat	taaacatgta	attgaaaac			876																								
<210>	40																													
<211>	230																													
<212>	PRT																													
<213>	Discosoma species																													
<400>	40																													
Met	Ser	Cys	Ser	Lys	Asn	Val	Ile	Lys	Glu	Phe	Met	Arg	Phe	Lys	Val															
				5					10					15																
Arg	Met	Glu	Gly	Thr	Val	Asn	Gly	His	Glu	Phe	Glu	Ile	Lys	Gly	Glu															
			20					25					30																	
Gly	Glu	Gly	Arg	Pro	Tyr	Glu	Gly	His	Cys	Ser	Val	Lys	Leu	Met	Val															
		35					40					45																		
Thr	Lys	Gly	Gly	Pro	Leu	Pro	Phe	Ala	Phe	Asp	Ile	Leu	Ser	Pro	Gln															
	50					55					60																			

-25-

Phe	Gln	Tyr	Gly	Ser	Lys	Val	Tyr	Val	Lys	His	Pro	Ala	Asp	Ile	Pro
65					70					75					80
Asp	Tyr	Lys	Lys	Leu	Ser	Phe	Pro	Glu	Gly	Phe	Lys	Trp	Glu	Arg	Val
				85					90					95	
Met	Asn	Phe	Glu	Asp	Gly	Gly	Val	Val	Thr	Val	Ser	Gln	Asp	Ser	Ser
			100					105					110		
Leu	Lys	Asp	Gly	Cys	Phe	Ile	Tyr	Glu	Val	Lys	Phe	Ile	Gly	Val	Asn
			115				120					125			
Phe	Pro	Ser	Asp	Gly	Pro	Val	Met	Gln	Arg	Arg	Thr	Arg	Gly	Trp	Glu
	130					135					140				
Ala	Ser	Ser	Glu	Arg	Leu	Tyr	Pro	Arg	Asp	Gly	Val	Leu	Lys	Gly	Asp
145					150					155					160
Ile	His	Met	Ala	Leu	Arg	Leu	Glu	Gly	Gly	Gly	His	Tyr	Leu	Val	Glu
				165					170					175	
Phe	Lys	Ser	Ile	Tyr	Met	Val	Lys	Lys	Pro	Ser	Val	Gln	Leu	Pro	Gly
			180					185					190		
Tyr	Tyr	Tyr	Val	Asp	Ser	Lys	Leu	Asp	Met	Thr	Ser	His	Asn	Glu	Asp
		195					200					205			
Tyr	Thr	Val	Val	Glu	Gln	Tyr	Glu	Lys	Thr	Gln	Gly	Arg	His	His	Pro
	210					215					220				
Phe	Ile	Lys	Pro	Leu	Gln										
225					230										

<210> 41  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> m-att;

<221> misc\_difference  
 <222> 18  
 <223> n is a or g or c or t/u

<400> 41  
 rkycwgcttt yktrtacnaa stsgb

25

<210> 42  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> m-attB;

<221> misc\_difference  
 <222> 18  
 <223> n is a or g or c or t/u

<400> 42  
 agccwgcttt yktrtacnaa ctsgb

25

<210> 43  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> m-attr

<221> misc\_difference  
 <222> 18  
 <223> n is a or g or c or t/u



<400> 43  
gttcagcttt cktrtacnaa ctsgb 25

<210> 44  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> m-attL

<221> misc\_difference  
<222> 18  
<223> n is a or g or c or t/u

<400> 44  
agccwgcttt cktrtacnaa gtsgb 25

<210> 45  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> m-attP1

<221> misc\_difference  
<222> 18  
<223> n is a or g or c or t/u

<400> 45  
gttcagcttt yktrtacnaa gtsgb 25

<210> 46  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attB1

<400> 46  
agcctgcttt tttgtacaaa cttgt 25

<210> 47  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attB2

<400> 47  
agcctgcttt cttgtacaaa cttgt 25

<210> 48  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attB3

<400> 48  
accagcttt cttgtacaaa cttgt 25

<210> 49

<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attR1

<400> 49  
gttcagcttt tttgtacaaa cttgt 25

<210> 50  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attR2

<400> 50  
gttcagcttt cttgtacaaa cttgt 25

<210> 51  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attR3

<400> 51  
gttcagcttt cttgtacaaa gttgg 25

<210> 52  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attL1

<400> 52  
agcctgcttt tttgtacaaa gttgg 25

<210> 53  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attL2

<400> 53  
agcctgcttt cttgtacaaa gttgg 25

<210> 54  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attL3

<400> 54  
acccagcttt cttgtacaaa gttgg 25

<210> 55  
<211> 25

<212> DNA  
<213> Artificial Sequence

<220>  
<223> attP1

<400> 55  
gttcagcttt tttgtacaaa gttgg 25

<210> 56  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> attP2,P3

<400> 56  
gttcagcttt cttgtacaaa gttgg 25

<210> 57  
<211> 34  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Lox P site

<400> 57  
ataacttcgt ataatgtatg ctatacgaag ttat 34

<210> 58  
<211> 1032  
<212> DNA  
<213> Escherichia coli

<220>  
<221> CDS  
<222> (1)...(1032)  
<223> nucleotide sequence encoding Cre recombinase

<400> 58  
atg tcc aat tta ctg acc gta cac caa aat ttg cct gca tta ccg gtc 48  
Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val  
1 5 10 15

gat gca acg agt gat gag gtt cgc aag aac ctg atg gac atg ttc agg 96  
Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg  
20 25 30

gat cgc cag gcg ttt tct gag cat acc tgg aaa atg ctt ctg tcc gtt 144  
Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val  
35 40 45

tgc cgg tcg tgg gcg gca tgg tgc aag ttg aat aac cgg aaa tgg ttt 192  
Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe  
50 55 60

ccc gca gaa cct gaa gat gtt cgc gat tat ctt cta tat ctt cag gcg 240  
Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala  
65 70 75 80

cgc ggt ctg gca gta aaa act atc cag caa cat ttg ggc cag cta aac 288  
Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn  
85 90 95

atg ctt cat cgt cgg tcc ggg ctg cca cga cca agt gac agc aat gct 336

Met	Leu	His	Arg	Arg	Ser	Gly	Leu	Pro	Arg	Pro	Ser	Asp	Ser	Asn	Ala		
			100					105					110				
ggt	tca	ctg	ggt	atg	cgg	cgg	atc	cga	aaa	gaa	aac	ggt	gat	gcc	ggt	384	
Val	Ser	Leu	Val	Met	Arg	Arg	Ile	Arg	Lys	Glu	Asn	Val	Asp	Ala	Gly		
		115					120					125					
gaa	cgt	gca	aaa	cag	gct	cta	gcg	ttc	gaa	cgc	act	gat	ttc	gac	cag	432	
Glu	Arg	Ala	Lys	Gln	Ala	Leu	Ala	Phe	Glu	Arg	Thr	Asp	Phe	Asp	Gln		
	130					135					140						
ggt	cgt	tca	ctc	atg	gaa	aat	agc	gat	cgc	tgc	cag	gat	ata	cgt	aat	480	
Val	Arg	Ser	Leu	Met	Glu	Asn	Ser	Asp	Arg	Cys	Gln	Asp	Ile	Arg	Asn		
	145				150					155					160		
ctg	gca	ttt	ctg	ggg	att	gct	tat	aac	acc	ctg	tta	cgt	ata	gcc	gaa	528	
Leu	Ala	Phe	Leu	Gly	Ile	Ala	Tyr	Asn	Thr	Leu	Leu	Arg	Ile	Ala	Glu		
			165						170					175			
att	gcc	agg	atc	agg	ggt	aaa	gat	atc	tca	cgt	act	gac	ggt	ggg	aga	576	
Ile	Ala	Arg	Ile	Arg	Val	Lys	Asp	Ile	Ser	Arg	Thr	Asp	Gly	Gly	Arg		
			180					185					190				
atg	tta	atc	cat	att	ggc	aga	acg	aaa	acg	ctg	ggt	agc	acc	gca	ggt	624	
Met	Leu	Ile	His	Ile	Gly	Arg	Thr	Lys	Thr	Leu	Val	Ser	Thr	Ala	Gly		
		195					200					205					
gta	gag	aag	gca	ctt	agc	ctg	ggg	gta	act	aaa	ctg	gtc	gag	cga	tgg	672	
Val	Glu	Lys	Ala	Leu	Ser	Gly	Val	Thr	Lys	Leu	Val	Glu	Arg	Trp			
	210					215					220						
att	tcc	gtc	tct	ggg	gta	gct	gat	gat	cgc	aat	aac	tac	ctg	ttt	tgc	720	
Ile	Ser	Val	Ser	Gly	Val	Ala	Asp	Asp	Pro	Asn	Asn	Tyr	Leu	Phe	Cys		
	225				230				235						240		
cgg	gtc	aga	aaa	aat	ggg	ggt	gcc	gcg	cca	tct	gcc	acc	agc	cag	cta	768	
Arg	Val	Arg	Lys	Asn	Gly	Val	Ala	Ala	Pro	Ser	Ala	Thr	Ser	Gln	Leu		
			245						250					255			
tca	act	cgc	gcc	ctg	gaa	ggg	att	ttt	gaa	gca	act	cat	cga	ttg	att	816	
Ser	Thr	Arg	Ala	Leu	Glu	Gly	Ile	Phe	Glu	Ala	Thr	His	Arg	Leu	Ile		
			260					265					270				
tac	ggc	gct	aag	gat	gac	tct	ggg	cag	aga	tac	ctg	gcc	tgg	tct	gga	864	
Tyr	Gly	Ala	Lys	Asp	Asp	Ser	Gly	Gln	Arg	Tyr	Leu	Ala	Trp	Ser	Gly		
		275					280					285					
cac	agt	gcc	cgt	gtc	gga	gcc	gcg	cga	gat	atg	gcc	cgc	gct	gga	ggt	912	
His	Ser	Ala	Arg	Val	Gly	Ala	Ala	Arg	Asp	Met	Ala	Arg	Ala	Gly	Val		
	290				295						300						
tca	ata	cgc	gag	atc	atg	caa	gct	ggg	ggc	tgg	acc	aat	gta	aat	att	960	
Ser	Ile	Pro	Glu	Ile	Met	Gln	Ala	Gly	Gly	Trp	Thr	Asn	Val	Asn	Ile		
	305				310					315					320		
gtc	atg	aac	tat	atc	cgt	aac	ctg	gat	agt	gaa	aca	ggg	gca	atg	gtg	1008	
Val	Met	Asn	Tyr	Ile	Arg	Asn	Leu	Asp	Ser	Glu	Thr	Gly	Ala	Met	Val		
				325					330					335			
cgc	ctg	ctg	gaa	gat	ggc	gat	tag									1032	
Arg	Leu	Leu	Glu	Asp	Gly	Asp	*										
			340														

<210> 59  
 <211> 343  
 <212> PRT

<213> Escherichia coli

<400> 59

```

Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val
1      5      10      15
Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg
      20      25      30
Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val
      35      40      45
Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe
      50      55      60
Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala
      65      70      75      80
Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn
      85      90      95
Met Leu His Arg Arg Ser Gly Leu Pro Arg Pro Ser Asp Ser Asn Ala
      100      105      110
Val Ser Leu Val Met Arg Arg Ile Arg Lys Glu Asn Val Asp Ala Gly
      115      120      125
Glu Arg Ala Lys Gln Ala Leu Ala Phe Glu Arg Thr Asp Phe Asp Gln
      130      135      140
Val Arg Ser Leu Met Glu Asn Ser Asp Arg Cys Gln Asp Ile Arg Asn
      145      150      155      160
Leu Ala Phe Leu Gly Ile Ala Tyr Asn Thr Leu Leu Arg Ile Ala Glu
      165      170      175
Ile Ala Arg Ile Arg Val Lys Asp Ile Ser Arg Thr Asp Gly Gly Arg
      180      185      190
Met Leu Ile His Ile Gly Arg Thr Lys Thr Leu Val Ser Thr Ala Gly
      195      200      205
Val Glu Lys Ala Leu Ser Leu Gly Val Thr Lys Leu Val Glu Arg Trp
      210      215      220
Ile Ser Val Ser Gly Val Ala Asp Asp Pro Asn Asn Tyr Leu Phe Cys
      225      230      235      240
Arg Val Arg Lys Asn Gly Val Ala Ala Pro Ser Ala Thr Ser Gln Leu
      245      250      255
Ser Thr Arg Ala Leu Glu Gly Ile Phe Glu Ala Thr His Arg Leu Ile
      260      265      270
Tyr Gly Ala Lys Asp Asp Ser Gly Gln Arg Tyr Leu Ala Trp Ser Gly
      275      280      285
His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val
      290      295      300
Ser Ile Pro Glu Ile Met Gln Ala Gly Gly Trp Thr Asn Val Asn Ile
      305      310      315      320
Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Thr Gly Ala Met Val
      325      330      335
Arg Leu Leu Glu Asp Gly Asp
      340

```

<210> 60

<211> 1272

<212> DNA

<213> Saccharomyces cerevisiae

<220>

<221> CDS

<222> (1)...(1272)

<223> nucleotide sequence encoding Flip recombinase

<400> 60

```

atg cca caa ttt ggt ata tta tgt aaa aca cca cct aag gtg ctt gtt
Met Pro Gln Phe Gly Ile Leu Cys Lys Thr Pro Pro Lys Val Leu Val
1      5      10      15
cgt cag ttt gtg gaa agg ttt gaa aga cct tca ggt gag aaa ata gca
Arg Gln Phe Val Glu Arg Phe Glu Arg Pro Ser Gly Glu Lys Ile Ala
      20      25      30

```

48

96

tta tgt gct gct gaa cta acc tat tta tgt tgg atg att aca cat aac	144
Leu Cys 35 Ala Ala Glu Leu Thr Tyr 40 Leu Cys Trp Met Ile Thr His Asn 45	
gga aca gca atc aag aga gcc aca ttc atg agc tat aat act atc ata	192
Gly Thr 50 Ala Ile Lys Arg Ala Thr Phe Met Ser Tyr 60 Asn Thr Ile Ile	
agc aat tcg ctg agt ttc gat att gtc aat aaa tca ctc cag ttt aaa	240
Ser Asn Ser Leu Ser Phe 70 Asp Ile Val Asn Lys 75 Ser Leu Gln Phe Lys 80	
tac aag acg caa aaa gca aca att ctg gaa gcc tca tta aag aaa ttg	288
Tyr Lys Thr Gln Lys 85 Ala Thr Ile Leu Glu Ala Ser Leu Lys Lys 95 Leu	
att cct gct tgg gaa ttt aca att att cct tac tat gga caa aaa cat	336
Ile Pro Ala Trp 100 Glu Phe Thr Ile Ile Pro Tyr Tyr Gly Gln Lys His 110	
caa tct gat atc act gat att gta agt agt ttg caa tta cag ttc gaa	384
Gln Ser Asp Ile Thr Asp Ile Val Ser Ser Leu Gln Leu Gln Phe Glu 125	
tca tcg gaa gaa gca gat aag gga aat agc cac agt aaa aaa atg ctt	432
Ser Ser 130 Glu Glu Ala Asp Lys 135 Gly Asn Ser His Ser Lys Lys Met Leu 140	
aaa gca ctt cta agt gag ggt gaa agc atc tgg gag atc act gag aaa	480
Lys Ala Leu Leu Ser 150 Glu Gly Glu Ser Ile Trp 155 Glu Ile Thr Glu Lys 160	
ata cta aat tcg ttt gag tat act tcg aga ttt aca aaa aca aaa act	528
Ile Leu Asn Ser Phe 165 Glu Tyr Thr Ser Arg Phe Thr Lys Thr Lys Thr 175	
tta tac caa ttc ctc ttc cta gct act ttc atc aat tgt gga aga ttc	576
Leu Tyr Gln Phe 180 Leu Phe Leu Ala Thr Phe Ile Asn Cys Gly Arg Phe 190	
agc gat att aag aac gtt gat ccg aaa tca ttt aaa tta gtc caa aat	624
Ser Asp Ile Lys Asn Val Asp 200 Lys Ser Phe Lys Leu Val Gln Asn 205	
aag tat ctg gga gta ata atc cag tgt tta gtg aca gag aca aag aca	672
Lys Tyr Leu Gly Val Ile Ile Gln Cys Leu Val Thr 220 Glu Thr Lys Thr 220	
agc gtt agt agg cac ata tac ttc ttt agc gca agg ggt agg atc gat	720
Ser Val Ser Arg His 230 Ile Tyr Phe Phe Ser Ala Arg Gly Arg Ile Asp 240	
cca ctt gta tat ttg gat gaa ttt ttg agg aat tct gaa cca gtc cta	768
Pro Leu Val Tyr 245 Leu Asp Glu Phe Leu Arg Asn Ser Glu Pro Val Leu 255	
aaa cga gta aat agg acc ggc aat tct tca agc aat aaa cag gaa tac	816
Lys Arg Val Asn Arg Thr Gly Asn Ser Ser Ser Asn Lys Gln Glu Tyr 270	
caa tta tta aaa gat aac tta gtc aga tcg tac aat aaa gct ttg aag	864
Gln Leu Leu Lys Asp Asn Leu Val Arg Ser Tyr Asn Lys Ala Leu Lys 285	
aaa aat gcg cct tat tca atc ttt gct ata aaa aat ggc cca aaa tct	912
Lys Asn Ala Pro Tyr Ser Ile Phe Ala Ile Lys 300 Gly Pro Lys Ser 290	

cac	att	gga	aga	cat	ttg	atg	acc	tca	ttt	ctt	tca	atg	aag	ggc	cta	960
His	Ile	Gly	Arg	His	Leu	Met	Thr	Ser	Phe	Leu	Ser	Met	Lys	Gly	Leu	
305					310					315					320	
acg	gag	ttg	act	aat	gtt	gtg	gga	aat	tgg	agc	gat	aag	cgt	gct	tct	1008
Thr	Glu	Leu	Thr	Asn	Val	Val	Gly	Asn	Trp	Ser	Asp	Lys	Arg	Ala	Ser	
				325					330					335		
gcc	gtg	gcc	agg	aca	acg	tat	act	cat	cag	ata	aca	gca	ata	cct	gat	1056
Ala	Val	Ala	Arg	Thr	Thr	Tyr	Thr	His	Gln	Ile	Thr	Ala	Ile	Pro	Asp	
			340					345					350			
cac	tac	ttc	gca	cta	gtt	tct	cgg	tac	tat	gca	tat	gat	cca	ata	tca	1104
His	Tyr	Phe	Ala	Leu	Val	Ser	Arg	Tyr	Tyr	Ala	Tyr	Asp	Pro	Ile	Ser	
		355					360					365				
aag	gaa	atg	ata	gca	ttg	aag	gat	gag	act	aat	cca	att	gag	gag	tgg	1152
Lys	Glu	Met	Ile	Ala	Leu	Lys	Asp	Glu	Thr	Asn	Pro	Ile	Glu	Glu	Trp	
	370					375					380					
cag	cat	ata	gaa	cag	cta	aag	ggg	agt	gct	gaa	gga	agc	ata	cga	tac	1200
Gln	His	Ile	Glu	Gln	Leu	Lys	Gly	Ser	Ala	Glu	Gly	Ser	Ile	Arg	Tyr	
385					390					395					400	
ccc	gca	tgg	aat	ggg	ata	ata	tca	cag	gag	gta	cta	gac	tac	ctt	tca	1248
Pro	Ala	Trp	Asn	Gly	Ile	Ile	Ser	Gln	Glu	Val	Leu	Asp	Tyr	Leu	Ser	
				405					410					415		
tcc	tac	ata	aat	aga	cgc	ata	taa									1272
Ser	Tyr	Ile	Asn	Arg	Arg	Ile	*									
			420													

<210> 61  
 <211> 422  
 <212> PRT  
 <213> *Saccharomyces cerevisiae*

<400> 61  
 Pro Gln Phe Gly Ile Leu Cys Lys Thr Pro Pro Lys Val Leu Val Arg  
 1 5 10 15  
 Gln Phe Val Glu Arg Phe Glu Arg Pro Ser Gly Glu Lys Ile Ala Leu  
 20 30  
 Cys Ala Ala Glu Leu Thr Tyr Leu Cys Trp Met Ile Thr His Asn Gly  
 35 40 45  
 Thr Ala Ile Lys Arg Ala Thr Phe Met Ser Tyr Asn Thr Ile Ile Ser  
 50 55 60  
 Asn Ser Leu Ser Phe Asp Ile Val Asn Lys Ser Leu Gln Phe Lys Tyr  
 65 70 75 80  
 Lys Thr Gln Lys Ala Thr Ile Leu Glu Ala Ser Leu Lys Lys Leu Ile  
 85 90 95  
 Pro Ala Trp Glu Phe Thr Ile Ile Pro Tyr Tyr Gly Gln Lys His Gln  
 100 105 110  
 Ser Asp Ile Thr Asp Ile Val Ser Ser Leu Gln Leu Gln Phe Glu Ser  
 115 120 125  
 Ser Glu Glu Ala Asp Lys Gly Asn Ser His Ser Lys Lys Met Leu Lys  
 130 135 140  
 Ala Leu Leu Ser Glu Gly Glu Ser Ile Trp Glu Ile Thr Glu Lys Ile  
 145 150 155 160  
 Leu Asn Ser Phe Glu Tyr Thr Ser Arg Phe Thr Lys Thr Lys Thr Leu  
 165 170 175  
 Tyr Gln Phe Leu Phe Leu Ala Thr Phe Ile Asn Cys Gly Arg Phe Ser  
 180 185 190  
 Asp Ile Lys Asn Val Asp Pro Lys Ser Phe Lys Leu Val Gln Asn Lys  
 195 200 205  
 Tyr Leu Gly Val Ile Ile Gln Cys Leu Val Thr Glu Thr Lys Thr Ser

210	Val	Ser	Arg	His	Ile	Tyr	215	Phe	Ser	Ala	Arg	220	Gly	Arg	Ile	Asp	Pro
225	Leu	Val	Tyr	Leu	Asp	Glu	230	Phe	Leu	Arg	Asn	235	Glu	Pro	Val	Leu	Lys
							245					250					
	Arg	Val	Asn	Arg	Thr	Gly	Asn	Ser	Ser	Ser	Asn	Lys	Gln	Glu	Tyr	Gln	
							260										
	Leu	Leu	Lys	Asp	Asn	Leu	Val	Arg	Ser	Tyr	Asn	Lys	Ala	Leu	Lys	Lys	
	Asn	Ala	Pro	Tyr	Ser	Ile	Phe	Ala	Ile	Lys	Asn	Gly	Pro	Lys	Ser	His	
							280										
	Ile	Gly	Arg	His	Leu	Met	Thr	Ser	Phe	Leu	Ser	Met	Lys	Gly	Leu	Thr	
305							295										
	Glu	Leu	Thr	Asn	Val	Gly	Asn	Trp	Ser	Asp	Lys	Arg	Ala	Ser	Ala		
	Val	Ala	Arg	Thr	Thr	Tyr	Thr	His	Gln	Ile	Thr	Ala	Ile	Pro	Asp	His	
	Tyr	Phe	Ala	Leu	Val	Ser	Arg	Tyr	Tyr	Ala	Tyr	Asp	Pro	Ile	Ser	Lys	
	Glu	Met	Ile	Ala	Leu	Lys	Asp	Glu	Thr	Asn	Pro	Ile	Glu	Glu	Trp	Gln	
	His	Ile	Glu	Gln	Leu	Lys	Gly	Ser	Ala	Glu	Gly	Ser	Ile	Arg	Tyr	Pro	
385																	
	Ala	Trp	Asn	Gly	Ile	Ile	Ser	Gln	Glu	Val	Leu	Asp	Tyr	Leu	Ser	Ser	
	Tyr	Ile	Asn	Arg	Arg	Ile											

<210> 62  
 <211> 48  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> IR2

<400> 62  
 gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttc

48

<210> 63  
 <211> 48  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> IR1

<400> 63  
 gaagttccta tactttctag agaataggaa cttcggaata ggaacttc

48

<210> 64  
 <211> 66  
 <212> DNA  
 <213> Bacteriophage mu

<220>  
 <221> CDS  
 <222> (1)...(66)  
 <223> nucleotide sequence encoding GIN recombinase

<400> 64  
 tca act ctg tat aaa aaa cac ccc gcg aaa cga gcg cat ata gaa aac  
 Ser Thr Leu Tyr Lys Lys His Pro Ala Lys Arg Ala His Ile Glu Asn  
 1 5 10 15

48

gac gat cga atc aat taa  
 Asp Asp Arg Ile Asn \*

66



20

<210> 65  
<211> 21  
<212> PRT  
<213> bacteriophage mu

<400> 65  
Ser Thr Leu Tyr Lys Lys His Pro Ala Lys Arg Ala His Ile Glu Asn  
1 5 10 15  
Asp Asp Arg Ile Asn  
20

<210> 66  
<211> 69  
<212> DNA  
<213> Bacteriophage mu

<220>  
<221> CDS  
<222> (1)...(69)  
<223> nucleotide sequence encoding Gin recombinase

<400> 66  
tat aaa aaa cat ccc gcg aaa cga acg cat ata gaa aac gac gat cga 48  
Tyr Lys Lys His Pro Ala Lys Arg Thr His Ile Glu Asn Asp Asp Arg  
1 5 10 15

atc aat caa atc gat cgg taa 69  
Ile Asn Gln Ile Asp Arg \*  
20

<210> 67  
<211> 22  
<212> PRT  
<213> bacteriophage mu

<220>  
<223> Gin recombinase of bacteriophage mu

<400> 67  
Tyr Lys Lys His Pro Ala Lys Arg Thr His Ile Glu Asn Asp Asp Arg  
1 5 10 15  
Ile Asn Gln Ile Asp Arg  
20

<210> 68  
<211> 555  
<212> DNA  
<213> Escherichia coli

<220>  
<221> CDS  
<222> (1)...(555)  
<223> nucleotide sequence encoding PIN recombinase

<400> 68  
atg ctt att ggc tat gta cgc gta tca aca aat gac cag aac aca gat 48  
Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp  
1 5 10 15

cta caa cgt aat gcg ctg aac tgt gca gga tgc gag ctg att ttt gaa 96  
Leu Gln Arg Asn Ala Leu Asn Cys Ala Gly Cys Glu Leu Ile Phe Glu  
20 25 30

gac aag ata agc ggc aca aag tcc gaa agg ccg gga ctg aaa aaa ctg	144
Asp Lys Ile Ser Gly Thr Lys Ser Glu Arg Pro Gly Leu Lys Lys Leu	
35 40 45	
ctc agg aca tta tgc gca ggt gac act ctg gtt gtc tgg aag ctg gat	192
Leu Arg Thr Leu Ser Ala Gly Asp Thr Leu Val Val Trp Lys Leu Asp	
50 55 60	
cgg ctg ggg cgt agt atg cgg cat ctt gtc gtg ctg gtg gag gag ttg	240
Arg Leu Gly Arg Ser Met Arg His Leu Val Val Leu Val Glu Glu Leu	
65 70 75 80	
cgc gaa cga ggc atc aac ttt cgt agt ctg acg gat tca att gat acc	288
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr	
85 90 95	
agc aca cca atg gga cgc ttt ttc ttt cat gtg atg ggt gcc ctg gct	336
Ser Thr Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala	
100 105 110	
gaa atg gag cgt gaa ctg att gtt gaa cga aca aaa gct gga ctg gaa	384
Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Lys Ala Gly Leu Glu	
115 120 125	
act gct cgt gca cag gga cga att ggt gga cgt cgt ccc aaa ctt aca	432
Thr Ala Arg Ala Gln Gly Arg Ile Gly Gly Arg Arg Pro Lys Leu Thr	
130 135 140	
cca gaa caa tgg gca caa gct gga cga tta att gca gca gga act cct	480
Pro Glu Gln Trp Ala Gln Ala Gly Arg Leu Ile Ala Ala Gly Thr Pro	
145 150 155 160	
cgc cag aag gtg gcg att atc tat gat gtt ggt gtg tca act ttg tat	528
Arg Gln Lys Val Ala Ile Ile Tyr Asp Val Gly Val Ser Thr Leu Tyr	
165 170 175	
aag agg ttt cct gca ggg gat aaa taa	555
Lys Arg Phe Pro Ala Gly Asp Lys *	
180	

<210> 69  
 <211> 184  
 <212> PRT  
 <213> Escherichia coli

<400> 69

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp	
1 5 10 15	
Leu Gln Arg Asn Ala Leu Asn Cys Ala Gly Cys Glu Leu Ile Phe Glu	
20 25 30	
Asp Lys Ile Ser Gly Thr Lys Ser Glu Arg Pro Gly Leu Lys Lys Leu	
35 40 45	
Leu Arg Thr Leu Ser Ala Gly Asp Thr Leu Val Val Trp Lys Leu Asp	
50 55 60	
Arg Leu Gly Arg Ser Met Arg His Leu Val Val Leu Val Glu Glu Leu	
65 70 75 80	
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr	
85 90 95	
Ser Thr Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala	
100 105 110	
Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Lys Ala Gly Leu Glu	
115 120 125	
Thr Ala Arg Ala Gln Gly Arg Ile Gly Gly Arg Arg Pro Lys Leu Thr	
130 135 140	
Pro Glu Gln Trp Ala Gln Ala Gly Arg Leu Ile Ala Ala Gly Thr Pro	
145 150 155 160	

Arg Gln Lys Val Ala Ile Ile Tyr Asp Val Gly Val Ser Thr Leu Tyr  
 165 170 175  
 Lys Arg Phe Pro Ala Gly Asp Lys  
 180

<210> 70  
 <211> 4778  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pcx plasmid

<400> 70  
 gtcgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60  
 gccccatataat ggaggttccgc gttacataaac ttacgggtaaa tggcccgccct ggctgaccgc 120  
 ccaacgacccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180  
 ggacttttcca ttgacgtcaa tgggtgggact atttacggta aactgcccac ttggcagtac 240  
 atcaagtgtg tcatatgcca agtacgcccc ctattgacgt caatgacggt aaatggcccg 300  
 cctggcatta tgcccagtac atgaccttat gggactttcc tacttggcag tacatctacg 360  
 tattagtcac cgctattacc atgggtcgag gtgagcccca cgttctgctt cactctcccc 420  
 atctcccccc cctccccacc cccaattttg tatttattta ttttttaatt attttgtgca 480  
 gcgatggggg cggggggggg gggggcgccg gccaggcggg gcggggcggg gcgagggggc 540  
 gggcgggggc aggcggagag gtgcggcgcc agccaatcag agcggcgcgc tccgaaagt 600  
 tccttttatg gcgagggcgcc ggcggcgccg gccctataaa aagcgaagcg cgcggcgggc 660  
 gggagtcgct gcgttgccct cgccccgtgc cccgctccgc gccgctcgc gccgcccgc 720  
 ccggtcttga ctgaccgctg tactcccaca ggtgagcggg cgggacggcc cttctcctcc 780  
 gggctgtaat tagcgcttgg ttaaatgacg gctcgtttct tttctgtggc tgcgtgaaag 840  
 ccttaaaggc ctccgggagg gccctttgtg cgggggggag cggctcgggg ggtgcgtgcg 900  
 tgtgtgtgtg cgtggggagg gccgcgtgcg cccgcgcgtg cccggcggtg gtgagcgctg 960  
 cgggcgcggc gcggggcctt gtgcgctccg cgtgtgcgcg aggggagcgc ggccgggggc 1020  
 ggtgccccgc ggtgcggggg ggctgcgagg ggaacaaagg ctgctgcgg ggtgtgtgcg 1080  
 tggggggggg agcagggggg gtgggcgcgg cggctcgggt gtaaccccc cctgcacccc 1140  
 cctccccgag ttgctgagca cggcccggtc tgggtgcgg ggctccgtgc ggggcgtggc 1200  
 gcgggggctc ccgtgcgggg cgggggggtg cggcaggtgg ggggtgcggg cggggcgggg 1260  
 ccgcctcggg ccggggaggg ctccggggag gccgcgcggc gggcgcgagc gccggcggt 1320  
 gtcgagggcg gcgagccgc agccattgcc ttttatggta atcgtgcgag agggcgagg 1380  
 gacttccttt gtcccaaatc tggcggagcc gaaatctggg aggcgcggcc gcaccccgc 1440  
 tagcggggcg gggcggaagc gtgcggcgcc ggcagggaag aaatggggcg ggagggcct 1500  
 cgtgcgtcgc gcgcccgcgc tccccttctc catctccagc ctgggggctg ccgcaggggg 1560  
 acggctgcct tcggggggga cggggcaggg cgggggttcg cttctggcgt gtgaccggcg 1620  
 gctctagagc ctctgctaac catgttcctt ccttcttctt tttcctacag ctctgggca 1680  
 acgtgctggt tgttgtgctg tctcatcatt ttggcaaga attcactcct caggtgcagg 1740  
 ctgcttatca gaagtggtg gctggtgtgg ccaatgccct ggctcacaaa taccactgag 1800  
 atctttttcc ctctgccaaa aattatgggg acatcatgaa gccccttgag catctgact 1860  
 ctggctaata aaggaaattt attttcattg caatagtgtg ttggaatttt ttgtgtctct 1920  
 cactcggaag gacatatggg agggcaaatc atttaaaaca tcagaatgag tatttggttt 1980  
 agagtgtggc aacatatgcc atatgctggc tgccatgaac aaaggtgggt ataaagaggt 2040  
 catcagtata tgaacagacc ccctgctgtc catccttat tccatagaaa agccttgact 2100  
 tgaggttaga ttttttttat attttgtttt gtgttatttt tttctttaac atccctaaaa 2160  
 ttttctttac atgttttact agccagattt ttctctctct cctgactact cccagtcata 2220  
 gctgtccctc ttctcttatg aagatccctc gacctgcagc ccaagcttgg cgtaatcatg 2280  
 gtcatagctg tttctgtgtt gaaattgtta tccgctcaca attccacaca acatacagac 2340  
 cggaagcata aagtgtaaag cctgggggtg ctaatgagtg agctaactca cattaattgc 2400  
 gttgcgctca ctgcccgcct tccagtcggg aaacctgtcg tgccagcgga tccgcatctc 2460  
 aattagtcag caaccatagt cccgccccct actccgcccc ttattttatg agagggcgag 2520  
 agttccgccc attctccgcc ccatggctga ctaatttttt tagtgaggag gcttttttgg 2580  
 gccgcctcgg cctctgagct attccagaag tagtgaggag gcttttttgg aggcctaggg 2640  
 ttttgcaaaa agctaaattg tttattgcag cttataatgg ttacaaataa agcaatagca 2700  
 tcacaaattt cacaaataaa gcattttttt cactgcattc tagttgtggt ttgtcccaac 2760  
 tcatcaatgt atcttatcat gctctggatcc gctgcattaa tgaatcgccc aacgcgcggg 2820  
 gagaggcggg ttgcgtattg ggcgctcttc cgcttccctc ctactgact cgctgcgctc 2880  
 ggtcgttcgg ctgcggcgag cgggtatcag tcaactcaag gcggtaatcc ggttatccac 2940  
 agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggcaggaa 3000  
 ccgtaaaaag gccgcgttgc tggcggtttt ccataggctc cgccccctg acgagcatca 3060  
 caaaaatcga cgctcaagtc agaggtggcg aaacccgaca ggactataaa gataccagg 3120  
 gtttccccct ggaagctccc tcgtgcgctc tcctgttccg accctgccgc ttaccggata 3180

cctgtccgcc	tttctccctt	cggggaagcgt	ggcgctttct	caatgctcac	gctgtaggta	3240
tctcagttcg	gtgtagggtcg	tctcgtccaa	gctgggctgt	gtgcacgaac	cccccgttca	3300
gccccgaccgc	tgcgccttat	cgggtaacta	tctgtcttgag	tccaacccgg	taagacacga	3360
cttatcgcca	ctggcagcag	ccactggtaa	caggattagc	agagcgaggt	atgtaggcgg	3420
tgctacagag	ttcttgaagt	ggtggcctaa	ctacggctac	actagaagga	cagtatttgg	3480
tatctgcgct	ctgctgaagc	cagttaacct	cggaaaaaga	gttggtagct	cttgatccgg	3540
caaaacaaacc	accgctggta	gcggtgggtt	ttttgtttgc	aagcagcaga	ttacgcgcag	3600
aaaaaaaagga	tctcaagaag	atcctttgat	cttttctacg	gggtctgacg	ctcagtgga	3660
cgaaaactca	cgttaaagga	ttttgggtcat	gagattatca	aaaaggatct	tcacctagat	3720
cctttttaaat	taaaaatgaa	gttttaaatc	aatctaaagt	atatatgagt	aaacttggtc	3780
tgacagttac	caatgcttaa	tcagtgaggc	acctatctca	gcgatctgtc	tatttcggtc	3840
atccatagtt	gcctgactcc	ccgtcgtgta	gataactacg	atacgggagg	gcttaccatc	3900
tggccccagt	gctgcaatga	taccgcgaga	cccacgctca	ccggctccag	atttatcagc	3960
aataaaccag	ccagccggaa	gggcccagcg	cagaagtggt	cctgcaactt	tatccgcctc	4020
catccagctc	attaattggt	gccgggaagc	tagagtaagt	agttcgccag	ttaatagttt	4080
gcgcaacgtt	gttgccattg	ctacaggcat	cgtgggtgtca	cgctcgtcgt	ttggtagggc	4140
ttcattcagc	tccggttccc	aacgatcaag	gcgagttaca	tgatccccc	tggtgtgcaa	4200
aaaagcgggt	agctccttcg	gtcctccgat	cgttgtcaga	agtaagttgg	ccgcagtggt	4260
atcactcatg	gttatggcag	cactgcataa	ttctcttact	gtcatgccat	ccgtaagatg	4320
cttttctgtg	actggtagt	actcaacc	gtcattctga	gaatagtgt	tgccgcgacc	4380
gagttgtctt	tgcccggcgt	caatacggga	taataccgcg	ccacatagca	gaactttaaa	4440
agtgtctcat	attggaaaac	gttcttcggg	gcgaaaaactc	tcaaggatct	taccgctggt	4500
gagatccagt	tcgatgtaac	ccactcgtgc	acccaactga	tcttcagcat	cttttacttt	4560
caccagcgtt	tctgggtgag	caaaaaacagg	aaggcaaaat	gccgcaaaaa	aggggaataag	4620
ggcgacacgg	aaatgttgaa	tactcatact	cttccttttt	caatattatt	gaagcattta	4680
tcagggttat	tgtctcatga	gcggatacat	atttgaatgt	atttagaaaa	ataaacaat	4740
aggggttccg	cgcacatttc	ccgaaaaagt	gccacctg			4778

<210> 71  
 <211> 5510  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pCXeGFP plasmid

<400> 71						
gtcgacattg	attattgact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gccatataat	ggagttccgc	gttacataac	ttacggtaaa	tggccgcct	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	attacggta	aactgccac	ttggcagta	240
atcaagtgt	tcataatgcca	agtaacgccc	ctattgacgt	caatgacgg	aaatggccc	300
cctggcatta	tgcccagta	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattagtc	cgctattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctcccc	420
atctccccc	cctccccc	cccaattttg	tattttatta	ttttttaatt	attttgtgca	480
gcgatggggg	cggggggggg	ggggggcgcg	gccagggcgg	gcggggcggg	gcgagggggc	540
ggggcgggcg	aggcgagag	gtgcccggcg	agccaatcag	agcggcgcg	tccgaaagt	600
tccttttatg	gcgagcgcg	ggcgggcgcg	gccctataaa	aagcgaagcg	cgcggcgggc	660
gggagtcgct	gcgttgccct	cgccccgtgc	cccgtccgc	gcgcctcgc	gccgcccgcc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtgagcggg	cgggacggcc	cttctcctcc	780
gggctgta	tagcgtttgg	tttaatgacg	gctcgtttct	ttctgtggc	tgctgaaag	840
ccttaaagg	ctccgggagg	gccctttgtg	cgggggggag	cggtcggggg	ggtgcgtgcg	900
tgtgtgtgtg	cgtggggagc	gccgcgtgcg	gcccgcgctg	cccggcggt	gtgagcgctg	960
cgggcgcgcg	gcggggcttt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggccgggggg	1020
gggtgccccg	gggtgcgggg	ggctgcgagg	ggaacaaagg	ctgcgtgcgg	ggtgtgtgcg	1080
tgggggggtg	agcagggggg	gtggggcgcg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cggcccggt	tcgggtgcgg	ggctccgtgc	ggggcgtagg	1200
gcgggggctcg	ccgtgcgggg	cgggggggtg	cggcaggtgg	gggtgcgggg	cggggcgggg	1260
ccgcctcggg	cgggggagg	ctcgggggag	gggcgcggcg	gccccggagc	gccggcggt	1320
gtcagagcgc	ggcgagcgc	agccattgcc	tttttggt	atcgtagcag	agggcgagag	1380
gacttccttt	gtcccaaatc	tggcggagcc	gaaatctggg	agggcgccgc	gcacccctc	1440
tagcggggcg	ggcggaagcg	gtgcggcgcc	ggcaggaagg	aaatggcgcg	ggagggcctt	1500
cgtgcgtcgc	cgcgcgcgcg	tcccctctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acggctgcct	ctcggggggg	cggggcgagg	cgggggtcgg	cttctggcgt	gtgacggcg	1620
gctctagagc	ctctgctaag	catgttcatt	ccttctcttt	tttctacag	ctcctgggca	1680
acgtgctggg	ttgtgtgctg	tctcatcatt	ttggcaaa	attcgccacc	atggtagca	1740
agggcgagga	gctgttcacc	gggggtggtg	ccatcctggt	cgagctggac	ggcgacgtaa	1800

acggccacaa	gttcagcgtg	tccggcgagg	gcgagggcga	tgccacctac	ggcaagctga	1860
ccctgaagtt	catctgcacc	accggcaagc	tgcccgtgcc	ctggcccacc	ctcgtgacca	1920
ccctgacctt	cggtgtgcag	tgcttcagcc	gctaccccga	ccacatgaag	cagcacgact	1980
tcttcaagtc	cgccatgccc	gaaggctacg	tccaggagcg	caccatcttc	ttcaaggacg	2040
acggcaacta	caagacccgc	gcccaggtga	agttcgaggg	cgacacccctg	gtgaaccgca	2100
tgcagctgaa	gggcatcgac	ttcaaggagg	acggcaacat	cctgggggcac	aagctggagt	2160
acaactacaa	cagccacaa	gtctatatca	tgcccgacaa	gcagaagaac	ggcatcaagg	2220
tgaacttcaa	gatccgccac	aacatcgagg	acggcagcgt	gcagctcgcc	gaccataacc	2280
agcagaacac	ccccatcggc	gacggccccg	tgctgctgcc	cgacaaccac	tacctgagca	2340
cccagtcggc	cctgagcaaa	gacccccaacg	agaagcgcg	tcacatggtc	ctgctggagt	2400
tgcgtgaccg	cgccgggac	actctcgga	tgagcagcgt	gtacaagtaa	gaattcactc	2460
ctcaggtgca	ggctgcctat	cagaagggtg	tggtctgggt	ggccaatgcc	ctggctcaca	2520
aataccactg	agatcttttt	ccctctgcca	aaaattatgg	ggacatcatg	aagccccctg	2580
agcatctgac	ttctggctaa	taaaggaaat	ttattttcat	tgcaatagtg	tggtggaaat	2640
ttttgtgtct	ctcactcgga	aggacatatg	ggagggcaaa	tcatttaaaa	catcagaatg	2700
agtatttggt	ttagagtttg	gcaacatatg	ccatatgctg	gctgccatga	acaaagggtg	2760
ctataaagag	gtcatcagta	tatgaaacag	ccccctgctg	tccattcctt	attccataga	2820
aaagccttga	cttgagggtta	gatttttttt	atattttgtt	ttgtgttatt	tttttcttta	2880
acatccctaa	aattttcctt	acatgtttta	ctagccagat	ttttctcctt	ctcctgacta	2940
ctccagtcga	tagctgtccc	tcttctctta	tgaagatccc	tcgacctgca	gccaagcctt	3000
ggcgtaatca	tggtcatagc	tgtttctctg	gtgaaattgt	tatccgctca	caattccaca	3060
caacatacga	gcccgaagca	taaagtgtaa	agcctggggg	gcctaataag	tgagctaact	3120
cacattaatt	cggttgcgtc	cactgcccgc	tttccagtcg	ggaaacctgt	cgtyccagcg	3180
gatccgcac	tcaattagtc	agcaaccata	gtcccgcgcc	taactccgcc	catcccgcgc	3240
ctaactccgc	ccagttccgc	ccatttctcg	ccccatggct	gactaatttt	ttttatttat	3300
gcagaggccg	agggcgcttc	ggcctctgag	ctattccaga	agtagtgagg	aggctttttt	3360
ggaggccctg	gcttttgcaa	aaagctaact	tgtttattgc	agcttataat	ggttacaaat	3420
aaagcaatag	catcacaat	ttcacaata	aagcattttt	ttcactgcat	tctagttgtg	3480
gtttgcctaa	actcatcaat	gtatcttctc	atgtctggat	ccgctgcatt	aatgaatcgg	3540
ccaacggcgg	gggagagggc	gtttgcgtat	tgggcgctct	tccgcttctt	cgctcactga	3600
ctcgctgcgc	tcgggtcggtc	ggctgcggcg	agcggtatca	gctcactcaa	aggcggtaat	3660
acggttatcc	acagaatcag	gggataacgc	gctggcggtt	ttccataggc	aaggccagca	3720
tgacgagcat	cacaaaaatc	gacgctcaag	tcagaggtgg	cgaaaaccga	tccgcccccc	3780
aagataccag	gcgtttcccc	ctggaagctc	cctcgctgcg	tctcctgttc	caggactata	3840
gcttaccgga	tacctgtccg	cctttctccc	ttcggggaag	gtggcgcttt	cgacctgtcc	3900
acgctgtagg	tatctcagtt	cggtgtaggt	cgttcgctcc	aagctgggct	ctcaatgctc	3960
accccccggt	cagcccgacc	gctgcgcctt	atccggtaac	tatcgtcttg	gtgtgcacga	4020
ggtaagacac	gacttatcgc	cactggcagc	agccactggg	aacaggatta	agtcacaacc	4080
gtatgtaggc	gggtgtacag	agttcttgaa	gtgggtggct	aactacggct	gcagagcgag	4140
gacagtattt	ggatatctgcg	ctctgttgaa	gccagttacc	ttcggaaaaa	acactagaag	4200
ctcttgatcc	ggcaaacaaa	ccaccgctgg	tagcggtggg	ttttttgttt	gagttggtag	4260
gattacgcgc	agaaaaaaag	gatctcaaga	agatcctttg	atcttttcta	gcaagcagca	4320
cgctcagtg	aacgaaaaat	cacgttaagg	gatttttggt	atcttttcta	cggggtctga	4380
cttcacctag	atccttttaa	attaaaaatg	aagttttaaa	atgagattat	caaaaaaggat	4440
gtaaacttgg	tctgacagtt	accaatgctt	aatcagtgag	tcaatctaaa	gtatatatga	4500
tctatttcgt	tcatccatag	ttgcctgact	ccccgtcggt	gcacctatct	cagcgatctg	4560
gggcttacca	tctggcccca	gtgctgcaat	gataccgcga	tagataacta	cgatacggga	4620
agatttatca	gcaataaacc	agccagccgg	aaggggccgag	gacccaacgct	caccggctcc	4680
tttatccgcc	tccatccagt	ctattaattg	ttgcccggaa	cgcaagaagt	gtagttcgcc	4740
agttaatagt	ttgcgcgaacg	ttgttgccat	tgctacagga	gctagagtaa	cacgctcgtc	4800
gtttggtatg	gcttcattca	gctccgggtc	ccaacgatca	atcgtgggtg	catgatcccc	4860
catggtgtgc	aaaaaaagcg	ttagctcctt	cggtcctccg	aggcgagtta	gaagtaagtt	4920
ggccgcagtg	ttatcactca	tggttatggc	agcactgcat	atcgttgtca	ctgtcatgcc	4980
atccgtaaga	tgcttttctg	tgactgggtg	gtactcaacc	aagtcatctt	gagaatagtg	5040
tatcgggcga	ccgagttgct	cttgcccggc	gtcaatacgg	gataatacgg	cgccacatag	5100
cagaacttta	aaagtgtctc	tcattggaaa	acgttcttcg	gggcgaaaaa	tctcaaggat	5160
cttaccgctg	ttgagatcca	gttcgatgta	accactcgt	gcacccaact	gatcttcagc	5220
atcttttaact	ttccaccagcg	tttctgggtg	agcaaaaaaca	ggaaggcaaa	atgccgcgaa	5280
aaagggatac	agggcgacac	ggaaatgttg	aatactcata	ctcttctttt	ttcaatatta	5340
ttgaagcatt	tatcagggtt	attgtctcat	gagcggatag	atatttgaat	gtatttagaa	5400
aaataaacia	ataggggttc	cgcgacacatt	tccccgaaaa	gtgccacctg		5460
						5510

<210> 72  
 <211> 282  
 <212> DNA  
 <213> Artificial Sequence

<220>  
<223> attp

<400> 72  
ccttgcgcta atgctctgtt acagggtcact aataccatct aagtagttga ttcatagtga 60  
ctgcatatgt tgtgtttttac agtattatgt agtctgtttt ttatgcaaaa tctaatttaa 120  
tataattgata tttatatcat tttacgtttc tcgttcagct tttttatact aagttggcat 180  
tataaaaaaag cattgcttat caatttgttg caacgaacag gtcactatca gtcaaaaataa 240  
aatcattatt tgatttcaat tttgtcccac tccctgcctc tg 282

<210> 73  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 73  
ggccccgtaa tgcagaagaa 20

<210> 74  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 74  
ggtttaaaagt gcgctcctcc aagaacgtca tc 32

<210> 75  
<211> 40  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer

<400> 75  
agatctagag ccgccgctac aggaacaggt ggtggcggcc 40

<210> 76  
<211> 37  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer 5PacSV40

<400> 76  
ctgttaatta actgtggaat gtgtgtcagt taggggtg 37

<210> 77  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer Antisense Zeo

<400> 77  
tgaacaggggt cacgtcgtcc 20

<210> 78  
<211> 24

<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer 5' HETS

<400> 78  
gggccgaaac gatctcaacc tatt 24

<210> 79  
<211> 19  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer 3' HETS

<400> 79  
cgcagcggcc ctctactc 19

<210> 80  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer 5BSD

<400> 80  
accatgaaaa catttaacat ttctcaaca 29

<210> 81  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer SV40polyA

<400> 81  
tttatttgtg aaatttgtga tgctattgc 29

<210> 82  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer 3BSP

<400> 82  
ttaatttcgg gtatatttga gtgga 25

<210> 83  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer EPO5XBA

<400> 83  
tatctagaat ggggggtgcac gaatgtcctg cc 32

<210> 84  
<211> 32  
<212> DNA

<213> Artificial Sequence

<220>

<223> Primer EPO3SBI

<400> 84

tacgtacgtc atctgtcccc tgtcctgcag gc

32

<210> 85

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer GENEPO3BSI

<400> 85

cgtacgtcat ctgtcccctg tcctgca

27

<210> 86

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer GENEPO5XBA

<400> 86

tctagaatgg ggggtgcacgg tgagtact

28

<210> 87

<211> 4862

<212> DNA

<213> Artificial Sequence

<220>

<223> pD2eGFP-1N plasmid from Clontech

<400> 87

tagttattaa	tagtaaatcaa	ttacgggggtc	attagttcat	agcccatata	tggagttccg	60
cggtacataa	cttacggtaa	atggcccgcc	tggctgaccg	cccaacgacc	cccgccatt	120
gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	180
atgggtggag	tattttacggg	aaactgcccc	cttggcagta	catcaagtgt	atcatatgcc	240
aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgcccagta	300
catgacctta	tgggactttc	ctacttggga	gtacatctac	gtattagtca	tcgctattac	360
catgggtgatg	cgggttttggc	agtacatcaa	tgggcgtgga	tagcgggttg	actcacgggg	420
atttccaagt	ctccacccca	ttgacgtcaa	tgggagtgtg	ttttggcacc	aaaatcaacg	480
ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaatggggc	gtaggcggtg	540
acgggtgggag	gtctatataa	gcagagctgg	tttagtgaac	cgtcagatcc	gctagcgtca	600
ccggactcag	atctcgagct	caagcttcga	attctgcagt	cgacgggtacc	gcgggcccgg	660
gatccaccgg	tcgccaccat	gggtgagcaag	ggcgaggagc	tgttcaccgg	ggtaggtgcc	720
atcctgggtcg	agctggacgg	cgacgtaaac	ggccacaagt	tcagcgtgtc	cggcaggggc	780
gagggcgatg	ccacctacgg	caagctgacc	ctgaagttca	tctgcaccac	cggcaagctg	840
ccggtgccct	ggcccaccct	cgtgaccacc	ctgacctacg	gcgtgcagtg	cttcagccgc	900
taccccgacc	acatgaagca	gcacgacttc	ttcaagtccg	ccatgcccga	aggctacgtc	960
caggagcgca	ccatcttctt	caaggacgac	ggcaactaca	agacccgcgc	cgaggtgaag	1020
ttcgagggcg	acaccctggg	gaaccgcatc	gagctgaagg	gcacgcgactt	caaggaggac	1080
ggcaacatcc	tggggcacaa	gctggagtac	aactacaaca	gccacaacgt	ctatatcatg	1140
gccgacaagc	agaagaacgg	catcaagggtg	aacttcaaga	tccgccacaa	catcgaggac	1200
ggcagcgtgc	agctcgccga	ccactaccag	cagaacaccc	ccatcgggcga	cggccccgtg	1260
ctgctgcccc	acaaccacta	cctgagcacc	cagtccgccc	tgagcaaaaga	ccccaacgag	1320
aagcgcgata	acatgggtcct	gctggagtcc	gtgaccgccc	ccgggatcac	tctcggcatg	1380
gacgagctgt	acaagaagct	tagccatggc	ttcccgcggg	aggtggagga	gcaggatgat	1440
ggcagcgtgc	ccatgtcttg	tgcccaggag	agcgggatgg	accgtcacc	tgacgctgtg	1500
gcttctgcta	ggatcaatgt	gtagatgcgc	ggccgcgact	ctagatcata	atcagccata	1560
ccacatttgt	agaggtttta	cttgctttta	aaaactccc	acacctcccc	ctgaacctga	1620
aacataaaat	gaatgcaatt	gttggtgtta	acttggttat	tgcagcttat	aatggttaca	1680



```

aataaaagcaa tagcatcaca aatttcacaa ataaagcatt tttttcactg cattctagtt 1740
gtggtttgtc caaacatcatc aatgtatctt aaggcgtaaa ttgtaagcgt taatatatttg 1800
ttaaatttcg cgttaaatttt ttgttaaate agctcatttt ttaaccaata ggccgaaatc 1860
ggcaaaatcc cttataaatc aaaagaatag accgagatag ggttgagtgt tgttccagtt 1920
tggaacaaga gtccactatt aaagaacgtg gactccaacg tcaaagggcg aaaaaccgtc 1980
tatcagggcg atggccact acgtgaacca tcaccctaata caagtttttt ggggtcgagg 2040
tgccgtaaaag cactaaatcg gaaccctaaa gggagcccc aaggtatgagc ttgacgggga 2100
aagccggcga acgtggcgag aaaggaaggg aagaaagcga aaggagcggg cgctagggcg 2160
ctggcaagtg tagcggtcac gctgcgcgta accaccacac ccgccgcgct taatgcgcgc 2220
ctacagggcg cgtcaggtgg cacttttcgg ggaaatgtgc gcggaacccc tatttgttta 2280
tttttctaaa tacattcaaa tatgtatccg tcatgagac aataaccctg ataaatgctt 2340
caataatatt gaaaaaggaa gagtcctgag gcggaagaa ccagctgtgg aatgtgtgtc 2400
agttaggggtg tggaaagtcc ccaggctccc cagcaggcag aagtatgcaa agctatgcat 2460
tcaattagtc agcaaccagg tgtggaagt cccagggctc agtatgcaa agaagtatgc 2520
aaagcatgca tctcaattag tcagcaacca tagtcccgcc cccagcaggc cccatccgcg 2580
ccctaactcc gccaggttcc gcccattctc cgccccatgg ctgactaatt ttttttattt 2640
atgcagggcg cgagcccgcc tggcctctgt agctattcca gaagtatgga ggaggctttt 2700
ttggaggcct aggtttttgc aaagatcgat caagagacag gatgaggatc gtttcgcgat 2760
attgaacaag atggattgca aatcggtcgc cgggcccgtt cgtgagacag gctattcggc 2820
tatgactggg cggttctttt cgcaggttct tctgatcccg cgtgttccg gctgtcagcg 2880
cagggggcgcc cgcggtatc aagggactgg tgcctgtccg gtgcccgtgaa tgaactgcaa 2940
gacgaggcag caggttctca ctgaagcggg tctccttgcc ttccttgcc agctgtgctc 3000
gacgttctca ctgaagcggg tctccttgcc ttccttgcc agctgtgctc 3060
ctcctgtcat cggctgcata cgcttgatcc gtaactcgat cgcgcgcagc gccgcagcgc 3120
gagcgagcac gatcctggat tgcgtgaccca tggcgatgcc ctgttgcga atatcatggg 3180
catcaggggc gaggatctcg cgcttttctg cacttcatga ctgttgcga atatcatggg 3240
gcgttttgct ggcgtgatat gctatcgccgc gatcggttgg cggaccgcta tcaggacata 3300
gtgcttttac gatcgccgc gagcgggact tttcgattcc cggttggtatg gaaaaatggc 3360
gagtttctct ttcgattcc cggttggtatg gaaacacgga acccgcgcta 3420
catcacgaga tccgggacgc accctagggg aaaaagacag ccagggttgg ttccttctct 3480
accctagggg gaggttaact ccaaggttgg cactctgtcg tccccaccc gcaggccctg 3540
tgacggcaat gggttcggtc acgcccgcgt cgtcgggcg aaaacttcat ttttaattta 3600
tgcgagccaa agattgattt atctcatgac aaagatctta accgctacca aactggcttc 3660
agattgattt atctcatgac aaagatctta accgctacca aactggcttc 3720
aaaagatcaa caaaaaaac ttccgaaggt cgtagttagg ccaccacttc agtggttacc 3780
caaaaaaac ttccgaaggt cgtagttagg ccaccacttc agtggttacc 3840
gacgatagtt ccagcttggg gcgaacgacc tcccgaaggg cagcagggag cctctgactt 3900
ccagcttggg gcgaacgacc tcccgaaggg cagcagggag cctctgactt 3960
ggtttcgcca tatggaaaaa ctacatggtt at

```

<210> 88  
 <211> 5192  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pIRESpuro2 plasmid from Clontech

<400> 88  
 gacggatcgg gagatctccc gatcccttat ggtcgactct cagtacaatc tgctctgatg 60  
 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gactagtgcg 120  
 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180

ttaggggttag	gogtttttgcg	ctgcttccgcg	atgtacggggc	cagatatagc	cgttgacatt	240
gattattgac	tagttatttaa	tagtaatcaa	ttacgggggtc	attagttcat	agcccatata	300
tggagttccg	cgttacataa	cttacggtaa	atggcccgcc	tggctgaccg	cccaacgacc	360
cccgccatt	gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	420
attgacgtca	atgggtggac	tatttacggg	aaactgcccc	cttggcagta	catcaagtgt	480
atcatatgcc	aagtacggcc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	540
atgccagta	catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	600
tcgctattac	catggtgatg	cggttttggc	agtacatcaa	tgggcgtgga	tagcgggttg	660
actcacgggg	atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	720
aaaatcaacg	ggactttcca	aaatgtcgta	acaactccgc	ccattgacg	caaattggcg	780
gtaggcgtgt	acgggtgggag	gtctatataa	gcagagctct	ctggctaact	agagaacca	840
ctgcttactg	gcttatcgaa	attaatacga	ctcactatag	ggagacccaa	gcttgggtacc	900
gagctcggat	cgatatctgc	ggcctagcta	gcgcttaagg	cctgttaacc	ggcgtacgt	960
ctccggattc	gaattcggat	ccgcggccgc	atagataact	gatccagtgt	gctggaatta	1020
attcgtctgt	tgcgagggcc	agctgttggg	ctgagtagtc	cctctcaaaa	gcgggcatga	1080
cttctcgctc	aagatttgtca	gtttccaaaa	acgaggagga	tttgatattc	acctggcccc	1140
cggtgatgcc	tttgaggggtg	gccgcgtcca	tctggtcaga	aaagacaatc	tttttgttgt	1200
caagcttgag	gtgtggcagg	cttgagatct	ggccatacac	ttgagtga	atgacatcca	1260
ctttgccttt	ctctccacag	gtgtccactc	ctgcaggtcg	ctgcaggtcg	agcatgcac	1320
tagggcggcc	aattccgccc	ctctccctcc	cccccccta	acgttactgg	ccgaagccgc	1380
ttggaataag	gccggtgtgc	gtttgtctat	atgtgatttt	ccaccatatt	gccgtctttt	1440
ggcaatgtga	gggcccggaa	acctggccct	gtcttcttga	cgagcattcc	taggggtcct	1500
tccctctcgc	ccaaagggaat	gcaaggctcg	ttgaatgtcg	tgaagggaagc	agttcctctg	1560
gaagcttctt	gaagacaaaac	aacgtctgtg	gcgacccttt	gcaggcagcg	gaacccccca	1620
cctggcgaca	ggtgcctctg	cgcccaaaa	ccacgtgtat	aagatacacc	tgcaaaaggcg	1680
gcacaacccc	agtggcacgt	tgtgagttgg	gcccagaagg	aaagagtcaa	atggctctcc	1740
tcaagcgtat	tcaacaagggg	gctgaaggat	tgtgtttagt	taccccatctg	tatgggatct	1800
gatctggggc	ctcggtgcac	atgctttaca	ctttgaaaaa	cgaggttaaa	aaaaacgtcta	1860
ggccccccga	accacgggga	cggtgttttc	cacgatgata	cacgatgata	agcttgccac	1920
aaccacacaag	gagacgacct	tccatgaccg	agtaacaagg	cacggtgcgc	ctcgccaccc	1980
gcgacgacgt	cccccgggcc	gtacgcaccc	tcgcgcgcgc	gttcgcgcgc	taccgcgcgc	2040
cgccccacac	cgtcgaccgc	gaccgccaca	tcgagcgggt	caccgagctg	caagaactct	2100
tcctcacgcg	cgtcgggctc	gacatcgcca	aggtgtgggt	cgcggaagac	ggcgccgcgc	2160
tggcggtctg	gaccacgcgc	gagagcgtcg	aagcgggggc	ggtgttcgcc	gagatcgggc	2220
cgcccatggc	cgagttagcg	ggttcccgcc	tgcccgcgca	gcaacagatg	gaaggcctcc	2280
tggcgccgca	ccggcccaag	gagcccgcg	ggttccctgg	caccgtcggc	gtctcgcccg	2340
accaccaggg	caagggtctg	ggcagcgccg	tcgtgtctcc	cggagtggag	gcggcgagac	2400
gcgcccgggt	gccccgcctc	ctggagacct	ccgcgcgcgc	caacctcccc	ttctacgagc	2460
ggctcggctt	caccgtcacc	gccgacgtcg	agtgcccgaa	ggaccgcgcg	acctgggtgca	2520
tgaccgcgaa	gccccggtgc	tgacgcgcgc	cccacgaccc	gcagcgcccg	accgaaaggga	2580
gcgcacgac	ccatggctcc	gaccggaagc	gaccggggcg	gccccgcgca	ccccgcaccc	2640
gcccccgagg	cccaccgact	ctagagctcg	ctgatcagcc	tcgactgtgc	cttctagtgt	2700
ccagccatct	gttgtttgccc	cctccccctg	gccttctctg	accctgggaag	gtgccactcc	2760
cactgtcctt	tcctaataaaa	atgaggaaat	tgcattcgat	tgtctgagta	ggtgtcattc	2820
tattctgggg	ggtgggggtg	ggcaggacag	caagggggag	gattgggaag	acaatagcag	2880
gcatgctggg	gatgcgggtg	gctctatggc	ttctgaggcg	gaaagaacca	gctgggggtc	2940
tagtgcatte	tagttgtggg	ttgtccaaac	tcattcaatgt	atcttatcat	gtctgtatcc	3000
cgctgacctc	tagctagagc	ttggcgtaat	catggtcata	gctgtttcct	gtgtgaaatt	3060
gttatccgct	cacaattcca	cacaacatac	gagccggaag	cataaagtgt	aaagcctggg	3120
gtgcctaagt	agtgagctaa	ctcacattaa	ttgcgttgcg	ctcactgccc	gctttccagt	3180
cgggaaacct	gtcgtgccag	ctgcattaat	gaatcgccca	acgcgcgggg	agaggcggtt	3240
tgcgtattgg	gcgctcttcc	gcttccctcg	tcactgactc	gctgcgctcg	gtcgttcggc	3300
tgcggcgagc	ggtatcagct	cactcaaaag	cggtataacg	gttatccaca	gaatcagggg	3360
ataacgcagg	aaagaacatg	tgagcaaaa	gccagcaaaa	ggccagggaac	cgtaaaaagg	3420
ccgcgttgct	ggcggttttc	cataggctcc	gccccctga	cgagcatcac	aaaaatcgac	3480
gctcaagtga	gaggtggcga	aaccgcagac	gactataaag	ataccaggcg	tttccccctg	3540
gaagctccct	cgtgcgctct	cctgttccga	ccctgccgct	taccggatac	ctgtccgcct	3600
ttctcccttc	gggaagcgtg	gcgctttctc	aatgctcagc	ctgtagggtat	ctcagttcgg	3660
tgtaggtcgt	tcgctccaag	ctgggctgtg	tgacgaacc	ccccgttcag	cccgaccgtc	3720
gcgccttatc	cggtaactat	cgtcttgagt	ccaacccggg	aagacacgac	ttatcgccac	3780
tggcagcagc	cactggtaaac	aggattagca	gagcagggtg	tgtaggcggt	gctacagagt	3840
tcttgaagtg	gtggcctaac	tacggctaca	ctagaaggac	agtatttggt	atctcgctct	3900
tgctgaagcc	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	aaacaaacca	3960
ccgctggtag	cgggtggtttt	tttgtttgca	agcagcagat	tacgcgcaga	aaaaaaggat	4020
ctcaagaaga	tcctttgatc	ttttctacgg	ggtctgacgc	tcagtggaaac	gaaaactcac	4080
gttaagggat	tttgggtcatg	agattatcaa	aaaggatctt	cacctagatc	cttttaaaatt	4140
aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttgggtct	gacagttacc	4200

aatgcttaaat	cagttaggca	cctatctcag	cgatctgtct	atttcgttca	tccatagttg	4260
cctgactccc	cgctcgtgtag	ataactacga	tacggggaggg	cttaccatct	ggccccagtg	4320
ctgcaatgat	accgcgagac	ccacgctcac	cggctccaga	tttatcagca	ataaaccagc	4380
cagccggaag	ggccgagcgc	agaagtgggc	ctgcaacttt	atccgcctcc	atccagtcta	4440
ttaattgttg	ccgggaagct	agagtaagta	gttcgccagt	taatagtttg	cgcaacggtg	4500
ttgccattgc	tacaggcatc	gtgggtgtcac	gctcgtcgtt	tgggtatggct	tcattcagct	4560
ccggttccca	acgatcaagg	cgagttacat	gatcccccat	gttggtgcaaa	aaagcgggta	4620
gctccttcgg	tcctccgac	gttggtcagaa	gtaagttggc	cgcagtgtta	tcactcatgg	4680
ttatggcagc	actgcataat	tctcttactg	tcatgccatc	cgtaagatgc	ttttctgtga	4740
ctgggtgagta	aatacgggat	aataccgcgc	cacatagcag	gcggcgaccg	agttgctctt	4800
gcccggcgctc	ttcttcgggg	cgaaaactct	caaggatctt	aactttaaaa	gtgctcatca	4860
ttggaaaaacg	cactcgtgca	cccaactgat	cttcagcatc	accgctgttg	agatccagtt	4920
cgatgtaacc	aaaaacagga	aggcaaaatg	ccgcaaaaaa	ttttactttc	accagcgttt	4980
ctgggtgagc	actcatactc	ttcctttttc	aataattattg	gggaataagg	gcgacacgga	5040
aatggtgaat	cggatacata	tttgaatgta	tttagaaaaa	aagcatttat	caggggttatt	5100
gtctcatgag	ccgaaaagtg	ccacctgacg	tc	taaacaataa	gggggttccgc	5160
gcacatttcc						5192

<210> 89  
 <211> 11182  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pAg1 Plasmid

<400> 89						
catgccaaacc	acaggggttcc	cctcggggatc	aaagtacttt	gatccaaccc	ctccgctgct	60
atagtgcagt	cggcttctga	cgttcagtg	agccgtcttc	tgaaaaacgac	atgtcgcaca	120
agtcctaagt	tacgcgacag	gctgcgcgcc	tgcccttttc	ctggcggttt	cttgtcgcgt	180
gttttagtcg	cataaagtag	aatacttgcc	actagaaccg	gagacattac	gccatgaaca	240
agagcgccgc	cgctggcctg	ctgggctatg	cccgcgtcag	caccgacgac	caggacttga	300
ccaaccaacg	ggccgaactg	cacgcggccg	gctgcacca	gctgttttcc	gagaagatca	360
ccgggaccag	gcgcgaccgc	ccggagctgg	ccaggatgct	tgaccaccta	cgccctggcg	420
acgttgtgac	agtgaccagg	ctagaccgcc	tgccccgcag	caccgcgac	ctactggaca	480
ttgccgagcg	catccaggag	gcccggcgccg	gcctgcgtag	cctggcagag	ccgtgggccc	540
acaccaccac	gcccggccggc	cgcattggtgt	tgaccgtgtt	cgccggcatt	gccgagttcg	600
agcgttccct	aatcatcgac	cgcacccgga	gcgggcgcga	ggccgccaag	gcccaggccg	660
tgaagttttg	cccccgccct	accctcacc	cggcacagat	cgcgacgc	cgcgagctga	720
tcgaccagga	aggccgcacc	gtgaaaagagg	cggctgcact	gcttggcggtg	catcgctcga	780
ccctgtaccg	cgcacttgag	cgcagcgagg	aagtgcgcgc	caccgaggcc	aggcggcgcg	840
gtgccttccg	tgaggacgca	ttgaccgagg	ccgacgccc	ggcgcccgcc	gagaatgaac	900
gccaaagagga	acaagcatga	aaccgcacca	ggacggccag	gacgaaccgt	ttttcattac	960
cgaagagatc	gaggcggaga	tgatcgcggc	cgggtacgtg	ttcgagccgc	ccgcgcacgt	1020
ctcaaacggtg	cggctgcatg	aaatcctggc	cggtttgtct	gatgccaaag	tgccggccctg	1080
gccggccagc	ttggccgctg	aagaaaaccga	gcgcgcgcgt	ctaaaaaggt	gatgtgtatt	1140
tgagtaaaac	agcttgccgtc	atgcggtcgc	tgcgtatatg	atgcgatgag	taataaaca	1200
aatacgcaag	gggaacgcat	gaagggtatc	gctgtactta	accagaaaagg	cggttcaggc	1260
aagacgacca	tcgcaaccca	tctagcccgc	gccctgcaac	tcgccggggc	cgatgttctg	1320
ttagtccatt	ccgatcccca	gggcagtgc	cgcgattggg	cgcccggtgcg	ggaagatcaa	1380
ccgctaaccg	ttgtcggcat	cgaccggccc	acgattgacc	gcgacgtgaa	ggccatccggc	1440
cggcgcgact	tcgtagtgat	cgacggagcg	cccaggccgc	cggacttggc	tgtgtccgcg	1500
atcaaggcag	ccgacttcgt	gctgatccg	gtgcagccaa	gcccttacga	catatgggcc	1560
accgccgacc	tggtggagct	ggttaagcag	cgcattgagg	tcacggatgg	aaggctacaa	1620
gcggcctttg	tcgtgtcgcg	ggcgatcaaa	ggcacgcgca	tcggcggtga	ggttgccgag	1680
gcgctggccg	ggtacgagct	gcccattctt	gagtcgccga	tcacgcagcg	cgtgagctac	1740
ccaggcactg	ccgcccgcgg	cacaaccggt	cttgaatcag	aacccgaggg	cgacgctgcc	1800
cgcgaggtcc	aggcgcgtggc	cgttgaaatt	aaatcaaaac	tcatttgagt	taatgaggta	1860
aagagaaaaa	gagcaaaaagc	acaaaacacgc	taagtgcggg	ccgtccgagc	gcacgcagca	1920
gcaaggctgc	aacgttggcc	agcctggcag	acacgccagc	catgaagcgg	gtcaactttc	1980
agttgccggc	ggaggatcac	accaagctga	agatgtacgc	ggtacgccaa	ggcaagacca	2040
ttaccgagct	gctatctgaa	tacatcgccg	agctaccaga	gtaaatgagc	aaatgaataa	2100
atgagttagt	gaatttttagc	gactaaagg	ggcgccatgg	aaaatcaaga	acaaccaggc	2160
accgacgccg	tggaatgcc	catgtgtgga	ggaacgggcg	gttggccagg	cgtaagcggc	2220
tgggttgtct	gcccggccctg	caatggcact	ggaaccccca	agcccaggga	atcggcggtg	2280
cggtcgcaaa	ccatccggcc	cggtaacaaat	cggcgccggc	ctgggtgatg	acctgggtga	2340
gaagttgaag	gcccgcgagg	ccgcccagcg	gcaacgcac	gaggcagaag	cacgccccgg	2400

tgaatcgtgg	caagcggccg	ctgatcgaat	ccgcaaagaa	tcccggcaac	cgccggcagc	2460
cggtgcgcg	tcgattagga	agccgcccaa	ggcgacagag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tcgcagcatc	atggacgtgg	ccgtttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	ggatgatccg	tacgagcttc	cagacgggca	2640
cgtagaggtt	tcgcagggc	cgccggccat	ggccagtggt	tgggattacg	acctgggtact	2700
gatggcgggt	tcccattctaa	ccgaatccat	gaaccgatac	cgggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagtctctgcc	ggcgagccga	2820
tggcggaag	cagaaagacg	acctggtaga	aacctgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cggcggcctg	gtgacgggat	ccgaggggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaacc	ggcgggccgg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggacgtgct	3060
gacgggttcac	cccgattact	ttttgatcga	tcccggcatc	ggcggttttc	tctaccgcct	3120
ggcacgccgc	gcccgcaggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gcccggagagt	tcaagaagt	ctgtttcacc	gtgcgcaagc	tgatcggttc	3240
aaatgacctg	ccggagtagc	atltgaagga	ggaggcgggg	caggctggcc	cgatcctagt	3300
catgcgctac	cgcaacctga	tcgaggcgga	agcatccgcg	ggttcctaata	gtacggagca	3360
gatcgatagg	caaatggccc	tagcagggga	aaaagggtcga	aaagggtctct	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaaaccctg	acattgggaa	3480
cccaaagccg	tacattggga	accgggtcaca	catgtaagt	actgatataa	aagagaaaaa	3540
aggcgatttt	tccgcttaaa	actcttataa	actcttataa	actcttataa	cccgctggc	3600
ctgtgcataa	ctgtctggcc	agcgcacagc	cgaagagctg	caaaaagcgc	ctacccttcg	3660
gtcgctgcgc	tccctacgcc	ccgcgccttc	gcgtcgccct	atcgccggccg	ctggccgctc	3720
aaaaatggct	ggcctaagcc	caggcaatct	accagggcgc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	cgccgcccac	atcaaggcac	cctgcctcgc	gcgtttcggt	gatgacgggtg	3840
aaaacctctg	acacatgcag	ctcccggaga	cggtcacagc	ttgtctgtaa	gcggatgccg	3900
ggagcagaca	agcccgctcag	ggcgcgtcag	cggtgtgtgg	cgggtgtcgg	ggcgagcca	3960
tgacccagtc	acgtagcgat	agcggagtgt	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcgggtg	tgaataaccg	cacagatgcg	taaggagaaa	4080
ataccgcac	agcgttctct	ccgcttctct	ctcactcaaa	tcgctgcgct	cggtcggtcg	4140
gctgcggcga	gcggtatcag	tgtagcaaaa	ggcggtataa	cggttatcca	cagaatcagg	4200
ggataaacgca	ggaaaagaaca	tgtagcaaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcgttg	gtcgctttt	tccataggtt	ccgccccct	gacgagcatc	acaaaaatcg	4320
acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	agataaccagg	cgtttcccc	4380
tggaaagctcc	ctcgtgcgct	ctcctgttcc	gacccgtccg	cttaccggat	acctgtccgc	4440
ctttctccct	tgggaagcgc	tggcgttttc	tcatagctca	cgctgtaggt	atctcagttc	4500
ggtgtagggtc	gttcgctcca	agctgggctg	tgtgcacgaa	ccccccgttc	agcccgaccg	4560
ctgcgcctta	tccggttaact	atcgtcttga	gtccaaccgc	gtaagacacg	acttatcgcc	4620
actggcagca	gcccgtggt	acaggatgga	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tggtggccta	actacggcta	cactagaagg	acagtatttg	gtatctgcgc	4740
tctgtgaag	ccagttacct	tcggaaaaag	agtgtgtagc	tcttgatccg	gcaaacaaaa	4800
caccgctgg	agcgttggt	tttttggtt	caagcagcag	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	gctcagtgga	acgaaaaactc	4920
acgttaaggg	attttggtca	tgcattctag	gtactaaaaa	aattcatcca	gtaaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	ccccgtaag	tcaaaaaata	gctcgacata	5040
ctgttcttcc	ccgatatacct	ccctgatcga	ccggacgcag	aaggcaatgt	cataccactt	5100
gtccgccttg	ccgcttctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgctg	tctcccaggt	cgcggtggga	aaagacaagt	tcctcttcgg	gcttttccgt	5220
ctttaaaaaa	tcatacagct	cgccgggatc	tttaaatgga	gtgtcttctt	cccagttttc	5280
gcaatccaca	tcggccagat	cgttattcag	taagtaatcc	aattcggcta	agcggctgtc	5340
taagctattc	gtatagggac	aatccgatat	gtcgatggag	tgaaagagcc	tgatgcactc	5400
cgatacacgc	tcgataatct	tttcagggtc	ttgttcatct	tcatactctt	ccgagcaaaag	5460
gacgccatcg	gcctcactca	tgagcagatt	gctccagcca	tcatgcggtt	caaagtgcag	5520
gaccttttga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	ccggttccac	5580
atcatagggtg	gtccctttat	accggctgtc	cgctattttt	aaatataggt	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgagcggtga	5700
tttttcgac	agttttttca	attccgggtga	tattctcatt	ttagccattt	attatttctt	5760
tcctcttttc	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tccttgcatt	ctaaaaacct	aaataaccaga	aaacagcttt	ttcaaagttg	5880
ttttcaaagt	tggcgtataa	catagtatcg	acggagccga	ttttgaaacc	gcggtgatca	5940
caggcagcaa	cggttactga	ctgttacaat	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaacc	cggcagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaaag	6060
ctgcccgcct	tacaacggct	ctcccgcgtga	cgccgtcccg	gactgatggg	ctgcctgtat	6120
cgagtgggtga	ttttgtcccg	agctgcccgtg	ggggagctg	ttggctgggt	gggtggcagga	6180
tatatgtgtg	tgtaaacaaa	ttgacgctta	gacaacttaa	taacacattg	cggacgtttt	6240
taatgtactg	aattaacgcc	gaattaatc	gggggatctg	gatttttagta	ctggattttg	6300
gttttaggaa	ttagaaattt	tattgataga	agtattttac	aaatacaaat	acataactaag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctatagg	aaccctaatt	cccttatctg	6420

ggaactactc	acacattatt	atggagaaac	tccagtgcaa	tctcgggtgac	gggcagggacc	6480
ggacggggcg	gtaccggcag	gctgaagtcc	agctgccaga	aacccacgtc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgccctcgtc	6600
atgcgcacgc	tccgggtcggt	gggcagcccgc	atgacagcga	ccacgctctt	gaagccctgt	6660
gcctccaggg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcccgt	ccgctgggtg	6720
cgggggggaga	cgtacacggg	cgactcggcc	gtccagtcgt	aggcgtttcg	tgccctccag	6780
gggcccgcgt	aggcgatgcc	ggcgaacctc	ccgtccacct	cggcgacgag	ccagggatag	6840
cgctcccgcg	gacggacgag	gtcgtccgtc	cactcctgcg	gttccctgcg	ctcggtagcg	6900
aagttgaccg	tgcttgtctc	gatgtagtgg	ttgacgatgg	tgacagccgc	cgccatgtcc	6960
gcctcgggtg	cacggcggat	gtcggccggg	cgtcgttctg	ggctcatggg	agactcgaga	7020
gagatagatt	tgtagagaga	gactgggtgat	ttcagcgtgt	cctctccaaa	tgaaatgaac	7080
ttcctttatat	agaggaaggt	cttgcgaaag	atagtgggat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaaacgt	cttctttttc	7200
cacgatgctc	ctcgtgggtg	gggggtccatc	tttggggacca	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttccctttat	cgcaatgatg	gcatttgtag	gtgcccacct	ccttttctac	7320
tgtccttttg	atgaagtgc	agatagctgg	gcaattggaat	ccgaggaggt	ttcccgatat	7380
taccctttgt	tgaaaagtct	caatagccct	ttgggtctct	gagactgtat	ctttgatatt	7440
cttggagtag	acgagagtgt	cgtgctccac	catgttatca	catcaatcca	cttgcttttg	7500
agacgtgggt	ggaacgtctt	ctttttccac	gatgctcctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcttgaac	gatagccttt	cctttatcgc	aatgatggca	7620
tttgtaggtg	ccaccttctc	tttctactgt	ccttttgatg	aagtgcacga	tagctgggca	7680
atggaatccg	aggaggtttc	ccgatattac	cctttgttga	aaagtctcaa	tagccctttg	7740
gtcttctgag	actgtatctt	tgatattctt	ggagtagacg	agagtgtcgt	gctccaccat	7800
gttggcaagc	tgctctagcc	aatacgcaaa	ccgctctccc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcacgacag	gtttcccgag	tggaaaagcg	gcagtgcgag	caacgcgaat	7920
aatgtgagtt	agctcactca	ttaggcacc	caggctttac	actttatgct	tccggctcgt	7980
atgttgtgtg	gaattgtgag	cggataacaa	tttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agccttgact	agaggggtcg	cggatatcac	acatgataag	atacattgat	8100
gagtttggag	aaaccacaac	tagaatgcag	tgaaaaaaat	gctttatttg	tgaaatttgt	8160
gatgctattg	ctttattttgt	aaccattata	agctgcaata	aacaagttgg	ggtgggcgaa	8220
gaactccagc	atgagatccc	cgcgctggag	gatcatccag	ccggcgctcc	ggaaaacgat	8280
tccgaagccc	aacctttcat	agaaggcgcc	ggtggaaatcg	aaatctcgta	gcacgtgtca	8340
gtcctgtctc	tcggccacga	agtgcacgca	gttgccggcc	gggtcgcgca	gggcgaactc	8400
ccgccccccac	ggctgtctcg	cgatctcggt	catggccggc	ccggaggcgt	cccgggaagt	8460
cgtggacacg	actcccgacc	actcggcgta	cagctcgtcc	aggcccgcca	cccacaccca	8520
ggccagggtg	ttgtccggca	ccacctggtc	ctggaccgag	ctgatgaaca	gggtcacgtc	8580
gtcccgagacc	acacccggcga	agtcgtcctc	cacgaagtc	cggggagaacc	cgagccgggtc	8640
ggtccagaag	tcgacccgctc	cgccgacgct	cgcgcgggtg	agcaccggaa	cggcactggg	8700
caactttggc	atggatccag	atttcgctca	agttagtata	aaaaagcagg	cttcaatcct	8760
gcaggaattc	gatcgacact	ctcgtctact	ccaagaatat	caaagataca	gtctcagaag	8820
accaaagggc	tattgagact	tttcaacaaa	gggtaatatc	gggaaacctc	ctcggattcc	8880
attgcccagc	tatctgtcac	ttcatcaaaa	ggacagtaga	aaagggaagt	ggcacctaca	8940
aatgccatca	ttgcgataaa	ggaaaaggcta	tcgttcaaga	tgccctctgcc	gacagtggtc	9000
ccaaagatgg	acccccaccc	acgaggagca	tcgtggaaaa	agaagacgtt	ccaaccacgt	9060
cttcaaagca	agtggattga	tgtgataaca	tggtggagca	cgacactctc	gtctactcca	9120
agaatatcaa	agatacagtc	tcagaagacc	aaagggtctat	tgagactttt	caacaaaggg	9180
taatattcgg	aaacctcctc	ggattccatt	gccagctatg	ctgtcacttc	atcaaaagga	9240
cagtagaaaa	ggaaggtggc	acctacaaat	gccatcattg	cgataaagga	aaggctatcg	9300
ttcaagatgc	ctctgccgac	agtggtccca	aagatggacc	cccacccacg	aggagcatcg	9360
tggaaaaaga	agacgttcca	accacgtctt	caaagcaagt	ggattgatgt	gatattctca	9420
ctgacgtaag	ggatgacgca	caatccact	atccttcgca	agaccttctt	ctatataagg	9480
aagttcattt	catttgagaga	ggacacgctg	aaatcaccag	tctctctcta	caaactctatc	9540
tctctcgagc	tttcgcagat	ccgggggggg	aatgagatat	gaaaaagcct	gaactcaccg	9600
cgacgtctgt	cgagaagtgt	ctgatcgaaa	agttcgacag	cgtctccgac	ctgatgcagc	9660
tctcggaggg	cgaagaatct	cgtgctttca	gcttcgatgt	aggagggcgt	ggatatgtcc	9720
tgccgggtaaa	tagctgcgcc	gatgggtttc	acaaagatcg	ttatgtttat	cggcactttg	9780
catcggccgc	gtcctccgatt	ccgggaagtgc	ttgacattgg	ggagtttagc	gagagcctga	9840
cctattgcat	ctcccgcggt	gcacagggtg	tcacgttgca	agacctgcct	gaaaccgaac	9900
tgcccgctgt	tctacaaccg	gtcgcggagg	ctatggatgc	gatcgctgcg	gccgatctta	9960
gccagacgag	cgggttcggc	ccattcggag	cgcgaaggat	cgggtcaatac	actacatggc	10020
gtgatttcat	atgcgcgatt	gctgatcccc	atgtgtatca	ctggcgaact	gtgatggacg	10080
acacgcgtcag	tgccgtccgtc	gcgcaggctc	tcgatgagct	gatgcttttg	gccgaggact	10140
ccccggaagt	ccggcacctc	gtgcacgcgg	atctcggctc	caacaatgtc	ctgacggaca	10200
atggccgcat	aacagcggtc	attgactgga	gcgaggcgat	gttcggggat	ttcccaatacg	10260
aggtcgccaa	catcttcttc	tggaggccgt	ggttggcttg	tatggagcag	cagacgcgct	10320
acttcgagcg	gaggcatccg	gagcttgacg	gatccgggag	actccgggag	tatatgtctc	10380
gcattgggtct	tgaccaactc	tatcagagct	tggttgacgg	caatttcgat	gatgcagctt	10440

gggcgagg	tcgatgcgac	gcaatcgtcc	gatccggagc	cgggactgtc	gggcgtagac	10500
aaatcgcccc	cagaagcgcg	gccgtctgga	ccgatggctg	tgtagaagta	ctcgccgata	10560
gtggaaacgg	acgccccagc	actcgtccga	gggcaaagaa	atagagtaga	tgccgaccgg	10620
atctgtcgat	cgacaagctc	gagtttctcc	ataataatgt	gtgagtagtt	cccagataag	10680
ggaattaggg	ttcctatagg	gtttcgctca	tgtgttgagc	atataagaaa	cccttagtat	10740
gtatttgtat	ttgtaaaata	cttctatcaa	taaaatttct	aatttcctaaa	acaaaaatcc	10800
agtactaaaa	tccagatccc	ccgaattaat	tcggcggttaa	ttcagatcaa	gcttggcact	10860
ggccgtcggt	ttacaacgct	gtgactggga	aaacccctggc	gttaccacaac	ttaatcgctt	10920
tgcagcacat	ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	ccgatcgccc	10980
ttcccaacag	ttgcgcagcc	tgaatggcga	atgctagagc	agcttgagct	tggatcagat	11040
tgtcgtttcc	cgccttcagt	ttaaactatc	agtgtttgac	aggatatatt	ggcgggttaa	11100
cctaagagaa	aagagcggtt	attagaataa	cggatatatta	aaagggcggt	aaaaggttta	11160
tccgttcgtc	catttgtatg	tg				11182

<210> 90  
 <211> 8428  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pCambia3300 Plasmid

<400> 90						
catgcccaacc	acagggttcc	cctcgggatc	aaagtacttt	gatccaaccc	ctccgctgct	60
atagtgcagt	cggcttctga	cgttcagtgc	agccgtcttc	tgaaaaacgac	atgtcgcaca	120
agtcctaagt	tacgcgacag	gctgcgcccc	tgcccttttc	ctggcggttt	cttgtcgctg	180
gttttagtcg	cataaagtag	aatacttgcg	actagaacccg	gagacattac	gccatgaaca	240
agagcgccgc	cgctggccctg	ctgggctatg	cccgcgtcag	caccgacgac	caggacttga	300
ccaaccacacg	ggccgaactg	cacgcggccg	gctgcaccaa	gctgttttcc	gagaagatca	360
ccggcaccag	gcgcgaccgc	ccggagctgg	ccaggatgct	tgaccaccta	cgccctggcg	420
acgtttgtgac	agtgaccagg	ctagaccgcc	tggcccgacg	caccgcgac	ctactggaca	480
ttgccgagcg	catccaggag	gccggcgccg	gccttgctag	cctggcagag	ccgtgggccc	540
acaccaccac	gccggccggc	cgcatgggtg	tgaccgtgtt	cgccggcatt	gccgagttcg	600
agcggttccct	aatcatcgac	cgcaaccgga	gcgggcgcga	ggccgccaag	gcccagggcg	660
tgaagtttgg	ccccgcctt	accctcacc	cggcacagat	cgcgacgcc	cgcgagctga	720
tcgaccagga	aggccgcacc	gtgaaagagg	cggctgcact	gcttggcggt	catcgctcga	780
ccctgtaccg	cgcacttgag	cgcagcgagg	aatgtacgcc	caccgaggcc	aggcggcgcg	840
gtgccttccg	ttgaccgaga	ccgagccctt	ggcggccgcc	ggcggccgcc	gagaatgaac	900
gccaagagga	acaagcatga	aaccgcacca	ggacggccag	gacgaaccgt	ttttcattac	960
cgaagagatc	gaggcggaga	tgatcgccgg	cgggtacgtg	ttcgagccgc	ccgcgcacgt	1020
ctcaaccgtg	cggctgcctg	aaatcctggc	cggtttgtct	gatgccaaag	tggcggcctg	1080
gccggccagc	ttggccgctg	aagaaaccga	gcgcgcgcgt	ctaaaaaggt	gatgtgtatt	1140
tgagttaaaac	agcttgcgtc	atgcggctgc	tgcttatatg	atgcgatgag	taaaataaaca	1200
aatacgcaag	gggaacgcac	gaagggttat	gctgtactta	accagaaagg	cgggtcaggg	1260
aagacgacca	tcgcaaccac	ctagccccgc	gccctgcaac	tcgcccgggg	cgatgttctg	1320
ttagtgcgatt	ccgatcccc	gggcagtgcc	cgcgattggg	cggccgtgcg	ggaagatcaa	1380
ccgctaaccg	ttgtcggcat	cgaccgcccc	acgattgacc	gcgacgtgaa	ggccatcgcc	1440
cggcgcgact	tcgtagtgat	cgacggagcg	cccaggcgcg	cggacttggc	tgtgtccgcg	1500
atcaaggcag	ccgacttcgt	gctgattccg	gtgcagccaa	gcccttacga	catatgggcc	1560
accgccgacc	tggtggagct	gggttaagcag	cgcattgagg	tcacggatgg	aaggctacaa	1620
gcggcctttg	tcgtgtcgcg	ggcgatcaaa	ggcacgcgca	tcggcggtga	ggttgcccag	1680
gcgctggccg	ggtacgagct	gccccattct	gagtcgccga	tcacgcagcg	cgtgagctac	1740
ccaggcactg	ccgccgcggg	cacaaccggt	cttgaatcag	aacccgaggg	cgacgctgcc	1800
cgcgaggtcc	aggcgctggc	cgttgaaatt	aaatcaaaac	tcatttgagt	taatgaggta	1860
aagagaaaat	gagcaaaaagc	acaaacacgc	taagtgcggg	ccgtccgagc	gcacgcagca	1920
gcaaggctgc	aacgttggcc	agcctggcag	acacgccagc	catgaagcgg	gtcaactttc	1980
agttgcccgg	ggaggatcac	accaagctga	agatgtacgc	ggtagcccaa	ggcaagacca	2040
ttaccgagct	gctatctgaa	tacatcgccg	agctaccaga	gtaaatgagc	aaatgaataa	2100
atgagtagat	gaatttttagc	ggctaaagga	ggcggcatgg	aaaatcaaga	acaaccaggc	2160
accgagcccg	tggatgtgga	catgtgtgga	ggcaacgggg	gttggccagg	cgtaagcgcc	2220
tgggttgtct	gccggccctg	caatggcact	ggaaccccc	agcccgagga	atcggcgtga	2280
cgggtcgcaa	ccatccggcc	cgttacaaat	cggcgccggc	ctgggtgatg	acctgggtga	2340
gaagttggaag	ccgcgcgagg	ccgccagcgc	gcaacgcac	gaggcagaag	cacgcccggg	2400
tgaatcgctg	caagcggccg	ctgatcgaat	ccgcaaagaa	tcccggcaac	cgccggcagc	2460
cgggtgcggc	tcgatttaga	agccgcacca	gggcgacgag	caaccagatt	ttttcgcttc	2520
gatgctctat	gacgtgggca	ccgcgcatag	tcgcagcatc	atggacgtgg	cggttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	gggtgatccgc	tacgagcttc	cagacgggca	2640



cgtagaggtt	tccgcagggc	cgcccgccat	ggccagtggt	tgggattacg	acctgggtact	2700
gatggcggtt	ttccatctaa	ccgaatccat	gaaccgatac	cggaagggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaag	cagaaagacg	acctggtaga	aacctgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cgccgccttg	gtgacggtat	ccgaggggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaaacc	gggcggccgg	agtacatcga	3000
gatcgagcta	gctgattgga	tgtaccgcga	gatcacagaa	ggcaagaacc	cgacgtgct	3060
gacggttcac	cccgtattact	ttttgatcga	tcccggcatc	ggccgttttc	tctaccgcct	3120
ggcacgccgc	gcccgcaggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtggcagc	gcccggagagt	tcaagaagtt	ctgtttcacc	gtgcgcaage	tgatcgggtc	3240
aatgacctg	ccggagtagc	atttgaagga	cattggggag	caggctggcc	cgatcctagt	3300
catgcgctac	cgcaacctga	tccgagggcga	agcatccgcc	ggttccctaat	gtacggagca	3360
gatgctaggg	caaattgccc	tagcagggga	aaaaggtcga	aaaggtctct	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaa	cggaaaccgt	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	catgtaagt	actgataata	aagagaaaaa	3540
aggcgatatt	tccgcctaaa	actctttaa	acttattaaa	actcttaaaa	cccgcctggc	3600
ctgtgcataa	ctgtctggcc	agcgcacagc	cattgggagc	caaaaaagcg	ctacccttcg	3660
gtcgctgcgc	tcctacgcgc	ccgcccgttc	gcgtcggcct	atcgccggccg	ctggccgcctc	3720
aaaaatggct	ggcctacggc	caggcaatct	accagggcgc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	ggcgccccc	atcaaggcac	cctggcctcg	gcgtttcggt	gatgcaggtg	3840
aaaacctctg	acacatgcag	ctcccggaga	cggtcacagc	ttgtctgtaa	gcggtatgccg	3900
ggagcagaca	agcccgctcag	ggcgcgctcag	cggtgtgttg	cggtgtcgg	ggcgagcca	3960
tgaccagtc	acgtagcgat	agcgaggtgt	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcggtg	tgaataaccg	cacagatcgc	taaggagaaa	4080
ataccgcac	agcgctctct	ccgcttctct	gctcactgac	tcgctgcgct	cggtcgttcg	4140
gctgcggcga	cggtatccag	ctcactcaaa	ggcggttaata	cggttatcca	cagaatcagg	4200
ggataaccga	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcgttg	ctggcggttt	tcctataggc	ccgccccctc	gacgagcatc	acaaaaatcg	4320
acgtccaagt	cagaggtggc	gaaacctgac	aggataataa	agataaccagg	cgtttccccc	4380
tggaagctcc	ctcgtgcgct	ctcctgttcc	gacctgccc	cttaccggat	acctgtccgc	4440
ctttctccct	tcgggaagcg	tggcgctttc	tcatagtcca	cgctgtaggt	atctcagttc	4500
gggttaggtc	gttctccca	agctgggtga	tgtgcagaa	cccccgctt	agcccgaccg	4560
ctgcgcctta	tcgggtaact	atcgtcttga	gtccaacccg	gtaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tggtggccta	actacggcta	acagtaagg	acagtatttg	gtatctgcgc	4740
tctgctgaag	ccagttacct	tcggaaaaag	agttggtagc	tcttgatccg	gcaaaacaa	4800
caccgctggg	agcggtgggt	tttttgtttg	caagcagcag	attacgcgca	gaaaaaaagc	4860
atctcaagaa	gatcctttga	tcctttctac	ggggtctgac	gctcagtga	acgaaaactc	4920
acgttaaggg	attttgggtc	tgcatcttag	gtactaaaa	aattcatcca	gtaaaaata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	5040
ctgttcttcc	ccgatattct	ccctgatcga	ccggagccag	aaggcaatgt	cataccactt	5100
gtccgcctcc	ccgcttctcc	caagatcaat	aaagccactt	actttgccat	ctttcacaaa	5160
gatgttgctg	tcctccagggt	cgccgtggga	aaagacaagt	tcctcttcgg	gcttttccgt	5220
ctttaaaaaa	tcatacagct	cgcgcgggat	tttaaatgga	gtgtcttctt	cccagttttc	5280
gcaatccaca	tcggccagat	cgttattcag	taagtaatcc	aattcggcta	agcggctgtc	5340
taagctattc	gtatagggac	aatccgatat	gtcgatggag	tgaaagagcc	tgatgcactc	5400
cgcatacagc	tcgataaatc	tttcagggtt	ttgttcactc	tcatactctt	ccgagcaaa	5460
gacgccatcg	gcctcactca	tgagcagatt	gctccagcca	tcatgccgtt	caaagtgcag	5520
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	ccggttccac	5580
atcatagggt	gtccctttat	accggctgtc	cgteattttt	aaatataggt	tttcattttc	5640
tcaccaccgc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgcagcggta	5700
tttttcgatc	agttttttca	attccgggtga	tattctcatt	ttagccattt	attatttctc	5760
tcctcttttc	tacagtattt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tccttgcat	ctaaaacctt	aaataccaga	aaacagcttt	ttcaaagtgt	5880
ttttcaaagt	tgccgtataa	catagtatcg	acggagccga	ttttgaaacc	gcgggtgatca	5940
caggcagcaa	cgctctgtca	tcgttacaat	caacatgcta	ccctccgcga	gatcatccgt	6000
gtttcaaacc	cgccagctta	gttgccgttc	ttccgaatag	catcggtaac	atgagcaaa	6060
tctgccgcct	tacaacggct	ctcccgtga	cgcctcccg	gactgatggg	ctgctgtat	6120
cgagtgggtg	ttttgtgccg	agctgcccgt	cggggagctg	ttggctggct	ggtggcagga	6180
tatatgtggg	tgtaaacaaa	ttgacgctta	gacaacttaa	taacacattg	cggacgtttt	6240
taatgtactg	aattaacgcc	gaattaatc	gggggatctg	gatttttagta	ctggattttg	6300
gttttaggaa	ttagaaattt	tattgataga	agttattttac	aaatacaaat	acataactag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accctatagg	aaccttaatt	cccttatctg	6420
ggaactactc	acacattatt	atggagaaac	tcgagtcaaa	tctcggtgac	gggcaggacc	6480
ggacggggcg	gtacggcgag	gctgaagtcc	agctgccaga	aaccacagtc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatc	ccgcgggggg	catatccgag	cgccctcgtc	6600
atgcgcacgc	tcgggtcggt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660

gcctccagg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcccg	ccgctgggtg	6720
cggggggaga	cgtacacgg	cgactcggcc	gtccagtcgt	aggcgttg	tgccctccag	6780
ggggcccggt	aggcgatgcc	ggcgaccccg	ccgtccaccc	cgggcgacgag	ccaggggata	6840
cgctcccgc	gacggacgag	gtcgtccg	cactccctg	gttccctg	ctcggta	6900
aaagttgac	tgcttgctc	gatgtagtgg	ttgacgatgg	tgacagaccg	cgcatgtcc	6960
gcctcggtgg	cacggcggt	gtcggccggg	cgctgtctc	ggctcatgg	agactcgaga	7020
gagatagatt	tgtagagaga	gactgggtg	ttcagcgtg	cctctccaaa	tgaaatgaac	7080
ttccttatat	agaggaaggt	cttgccaagg	atagtgggat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaacgt	cttctttttc	7200
cacgatgctc	ctcgtgggtg	gggggtccat	tttgggacca	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttcttttat	cgcaatgatg	gcattttgtag	gtgccacctt	cctttttctac	7320
tgctcttttg	atgaagtgc	agatagctgg	gcaatgggaat	ccgaggagggt	ttcccgatat	7380
taccctttgt	tgaaaagtct	caatagccct	ttggctcttct	gagactgtat	ccttgatat	7440
cttggagtag	acgagagtg	cgtgctccac	catgttatca	catcaatcca	cttgctttga	7500
agacgtgggt	ggaacgtctt	ctttttccac	gatgctcctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	ctcggcagag	catcttgaac	gatagccttt	cctttatcgc	aatgatggca	7620
ttttaggtg	ccaccttctt	tttctactgt	ccttttgatg	aaagtacaga	tagctgggca	7680
atggaatccg	aggaggtttc	ccgatattac	cctttgttga	aaagtctcaa	tagccctttg	7740
gtcttctgag	actgtatctt	tgatattctt	ggagtacag	agagtgtcgt	gctccaccat	7800
gttggcaagc	tgctctagcc	aatacgcgaa	ccgctctcc	ccgctgcttg	gccgattcat	7860
taatgcagct	ggcacgacag	gtttcccgac	tggaaagcgg	gcagtgcagc	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcaccc	caggctttac	actttatgct	tccggctcgt	7980
atgttggtg	gaattgtgag	cggaatacaa	ttcacacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agctcgggtac	ccggggatcc	tctagagtcg	acctgcaggc	atgcaagctt	8100
ggcactggcc	gtcgttttac	aacgtcgtga	ctgggaaaa	cctggcggtta	cccaacttaa	8160
tcgcttgca	gcacatcccc	ctttcgccag	ctggcgtaat	agcgaagagg	cccgcaacga	8220
tcgccccttc	caacagttgc	gcagcctgaa	tggcgaaatg	tagagcagct	tgagcttggg	8280
tcagattgtc	gtttcccgcc	ttcagtttaa	actatcagtg	tttgacagga	tatatggcg	8340
ggtaaaccta	agagaaaaga	gcgtttatta	gaataacgga	tatttaaaag	ggcgtgaaaa	8400
ggtttatccg	ttcgtccatt	tgtatgtg				8428

<210> 91

<211> 3438

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attBZeo Plasmid

<400> 91

tcgaccctct	agtcaaggcc	ttaagtgagt	cgtattacgg	actggccg	gttttacaac	60
gtcgtgactg	ggaaaaccct	ggcgttaccc	aacttaatcg	ccttgacgca	catccccctt	120
tcgccagctg	gcgtaatagc	gaagaggccc	gcaccgatcg	cccttcccaa	cagttgcgca	180
gcctgaatgc	cgaatggcgc	ttcgcttggt	aataaagccc	gcttcggcgg	gctttttttt	240
gttaactacg	tcaggtggca	cttttcgggg	aaatgtgcgc	ggaaccctta	tttgtttatt	300
tttctaaata	cattcaaata	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	360
ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	420
ttttgcggca	ttttgccttc	ctgtttttgc	tcacccagaa	acgctgggtg	aagtaaaaga	480
tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	acagcggtaa	540
gaccccttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagttct	600
gctatgtggc	gcggtattat	ccggtgttga	cgccggggca	gagcaactcg	gtcgccgcat	660
acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	720
tggcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	atgagtgata	acactgcggc	780
caacttactt	ctgacaacga	toggaggacc	gaaggagcta	accgcttttt	tgacacaacat	840
gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	ccataccaaa	900
cgacgagcgt	gacaccacga	tgccctgtagc	aatggcaaca	acgttcgcga	aactattaac	960
tggcgaaacta	cttactctag	cttcccgga	acaatttaata	gactggatgg	aggcggataa	1020
agttgcagga	ccacttctgc	gctcggccct	tccggctggc	tggtttattg	ctgataaatc	1080
tggagccgggt	gagcgtgggt	ctcgcgggat	cattgcagca	ctggggccag	atggtaagcc	1140
ctcccgatc	gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	1200
acagatcgct	gagataggtg	cctcactgat	taagcattgg	taactgtcag	accaagttta	1260
ctcatatata	ccttagattg	atttaacccg	gttgataatc	agaaaagccc	caaaaacagg	1320
aaagattgat	ttaagattat	ttaaattgta	aacgttaata	ttttgttaaa	attcgcgtaa	1380
aatttttgtt	aatcagctc	attttttaac	caataggccg	aaatcggcaa	aatcccttat	1440
aatcaaaaag	aatagcccca	gatagggttg	agtggtgttc	cagttttgga	caagagttca	1500
ctattaaaga	acgtggactc	caacgtcaaa	gggcgaaaaa	ccgtctatca	gggcgatggc	1560
ccactacgtg	aaccatcacc	caaatcaagt	tttttggggg	cgagggtgccg	taaagcacta	1620



```
aatcggaacc ctaaagggag ccccgattt agagcttgac ggggaaagcg aacgtggcga 1680
gaaaggaagg gaagaaagcg aaaggagcgg gcgctagggc gctggcaagt gtagcggtca 1740
cgctgcgcgt aaccaccaca cccgcgcgcg ttaatgcgcc gctacagggc gcgtaaaagg 1800
atctaggtga agatcctttt tgataatctc atgaccaaaa tcccttaacg tgagttttcg 1860
ttccactgag cgtcagaccc cgtagaaaag atcaaaggat cttcttgaga tccttttttt 1920
ctgcgcgtaa tctgctgctt gcaaacaaaa aaaccaccgc taccagcggg ggtttgtttg 1980
ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag agcgagata 2040
ccaaatactg ttcttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca 2100
ccgcctacat acctcgctct gctaatcctg ttaccagtgg ctgctgccag tggcgataag 2160
tcgtgtctta ccgggttggg ctcaagacga tagttaccgg ataaggcgca gcggtcgggc 2220
tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 2280
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aaggagaaaa ggcggacagg 2340
tatccggtaa gcggcagggg cggaacagga gagcgacga gggagcttcc agggggaaac 2400
gcctggatc tttatagtc ttgtcggttt cgccacctct gacttgagcg tcgatttttt 2460
tgatgctcgt cagggggggc gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg 2520
ttcctggcct tttgtggcc ttttgcctac atgtaatgtg agttagctca ctcataggc 2580
accccgaggt ttacacttta tgcctccggc tcgtatgttg tgtggaattg tgagcgata 2640
acaatttcac acaggaaaca gctatgacca tgattacgcc aagctacgta atacgactca 2700
ctagtggggc ccgtgcaatt gaagccggct ggcgccaaagc ttctctgcag gattgaagcc 2760
tgctttttta tactaacttg agcgaaatct ggatccatgg ccaagttgac cagtgcctt 2820
ccggtgctca ccgcgcgcga cgtcgcgcga gcggtcagat tctggaccga ccggtcggg 2880
ttctcccggg acttcgtgga ggacgacttc gccggtgtgg tccgggacga cgtgacctg 2940
ttcatcagcg ccgtccagga ccaggtggtg ccggaacaac ccctggcctg ggtgtgggtg 3000
cgcgccctgg acgagctgta cgccgagtg cgagatcgcc tccgaggtcg tgtccacgaa cttccgggac 3060
gcctccgggg ccggccatgac cgagatcgcc gagcagccgt gggggcgggg gttcgccctg 3120
cgcgaccgg cgcgcaactg cgtgcacttc gtggccgagg agcaggactg acacgtgcta 3180
cgagatttcg attccaccgc cgccttctat gaaaggttgg gcttcggaat cgttttccgg 3240
gacgcgggt ggatgatcct ccagcgcggg gatctcatgc tggagtctt cgtccacccc 3300
aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 3360
aataaagcat ttttttctat gcattctagt tgtggtttgt ccaactcat caatgtatct 3420
tatcatgtct gtataccg 3438
```

<210> 92  
<211> 10549  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> pCambia1302 Plasmid

<300>  
<308> Genbank #AF234398  
<309> 2000-04-24

```
<400> 92
catggtagat ctgactagta aaggagaaga acttttctact ggagttgtcc caattcttgt 60
tgaattagat ggtgatgtta atgggcacaa attttctgtc agtggagagg gtgaaggtga 120
tgcaacatac ggaaaactta cctttaaatt tatttgcact actggaaaac tacctgttcc 180
gtggccaaca cttgtcacta ctttctctta tgggtgttcaa tgcttttcaa gatacccaga 240
tcatatgaag cggcacgact tcttcaagag cgccatgcct gagggatacg tgcaggagag 300
gaccatcttc ttcaaggacg acgggaacta caagacacgt gctgaagtca agtttgaggg 360
agacaccctc gtcaacagga tctgagcttaa gggaatcgat ttcaaggagg acggaaacat 420
cctcgcccac aagttggaat acaactacaa ctcccacaaac gtatacatca tggccgacaa 480
gcaaaagaac ggcatacaag ccaacttcaa gaccgcgcac aacatcgaag acggcggtcg 540
gcaactcgct gatcattatc aacaaaatac tccaattggc gatggccctg tccttttacc 600
agacaaccat tacctgtcca cacaatctgc cctttcgaaa gatcccaacg aaaagagaga 660
ccacatggte cttcttgagt ttgtaacagc tgctgggatt acacatggca tggatgaact 720
atacaaagct agccaccacc accaccacca cgtgtgaatt ggtgaccagc tcgaatttcc 780
ccgatcgttc aaacatttgg caataaagtt tcttaagatt gaatcctgtt gcgggtcttg 840
cgatgattat catataattt ctggtgaatt aggttaagca tgtaataatt aacagtgtta 900
gcatgacgtt atttatgaga tgggttttta tgattagagt cccgcaatta tacatttaat 960
acgcgataga aaacaaaaata tagcgcgcaa actaggataa attatcgcgc gcggtgtcat 1020
ctatgttact atagatcgga ttaaactatc agtgtttgac aggatataat ggcgggtctaa 1080
cctaagagaa aagagcgttt attagaataa cggatattta aaagggcggtg aaaaggttta 1140
tccggttcgtc cctttgctatg tgcatgccaa ccacagggtt cccctcgggg tcaaagctact 1200
ttgatccaac caatccgctg ctatagtcca gctggttctt gacgttcagt gcagcgctct 1260
tctgaaaacg acatgtcgca caagtcctaa gttacgcgac aggtctgcgc cctgcccttt 1320
```

tcctggcggtt	ttctttgtcgc	gtgtttttagt	cgcataaagt	agaataacttg	cgactagaac	1380
cggagacatt	acgccatgaa	caagagcgcc	gccgctggcc	tgctgggcta	tgcccccgctc	1440
agcaccgacg	accaggactt	gaccaacca	cgggcccgaac	tgacgcggc	cggttgaccc	1500
aagctgtttt	ccgagaagat	caccggcacc	aggcgcgacc	gcccggagct	ggccaggatg	1560
cttgaccacc	tacgccctgg	cgacgttgtg	acagtgacca	ggctagaccg	ccctggcccg	1620
agcaccgcg	acctactgga	cattgcccag	cgcacccagg	aggccggcgc	gggcctgcgt	1680
agcctggcag	agccgtgggc	cgacaccacc	acgcggcgcc	gccgcattggt	gttgaccgtg	1740
ttcgccggca	ttgcccagtt	cgagcggttc	ctaactcatcg	accgcaccgc	gagcggggcgc	1800
gaggccgcca	aggcccgagg	cgtgaagttt	ggcccccgcc	ctaccctcac	ccccggcacag	1860
atcgcgacg	cccgcgagct	gatcgaccag	gaaggccgca	ccgtgaaaga	ggcggttgca	1920
ctgcttggcg	tgcatcgctc	gacctgttac	cgcgcaactt	agcgacggca	ggaagtgcag	1980
cccaccgagg	ccaggcgggc	cgggtgccttc	cgtgaggacg	cattgaccga	ggccgacgcc	2040
ctggcgggcg	ccgagaatga	acgccaagag	gaacaagcat	gaaaccgcac	caggacggcc	2100
aggacgaacc	gtttttcatt	accgaagaga	tcgaggcgga	gatgatcgcg	gccgggtacg	2160
tggtcgagcc	gcccgcgcac	gtctcaaccg	tgcggttgca	tgaatcctcg	gccggtttgt	2220
ctgatgccaa	gctggcgggc	tgcccgggca	gcttgggcgc	tgaagaaacc	gagcgccgcg	2280
gtctaaaaag	tttgatgtgt	tttgagttaa	acagcttgcg	tcattgcggtc	gctgcgtata	2340
tgatgcatg	agtaataaaa	caaatacgca	aggggaacgc	atgaagggtta	tcgctgtact	2400
taaccagaaa	ggcggttcag	gcaagacgac	catcgcaacc	catctagccc	gcgcccgtgca	2460
actcgccggg	ggcggttcag	tggttagtgc	ttccgatacc	cagggcagtg	cccgggattg	2520
ggcgggcggt	ggggaagatc	aaccgctaac	cgttgtcgcc	atcgaccgcc	cgacgattga	2580
cccgcgagct	aaggccatcg	gccggcgcgca	cttcgtagt	atcgacggag	cgccccaggc	2640
ggcggaactg	gctgtgtccg	cgatcaaggc	agccgacttc	gtgctgattc	cggtgcagcc	2700
aagcccttac	gacatatggg	ccaccgcccga	cctgggtggag	ctgggttaagc	agcgcatatga	2760
ggtcacggat	ggaaggctac	aagcgccctt	tgctgtgtcg	cgggcgatca	aaggcacgcg	2820
catcgccggt	gaggttgccg	agggcgctggc	cggttacgag	ctgcccattc	ttgagtcgcc	2880
tatcacgcag	cgctgagctg	acccaggcac	tgccgcccgc	ggcacaaccg	ttcttgaatc	2940
agaaccgcag	ggcgacgctg	cccgcgaggt	ccaggcgctg	gccgctgaaa	ttaaatcaaa	3000
actcatttga	gttaatgagg	taaagagaaa	atgagcaaaa	gcacaacac	gctaagtgc	3060
ggccgtccga	gcgacgcag	cagcaaggct	gcaacggttg	ccagcctggc	agacacgcca	3120
gccatgaagc	gggtcaactt	tcagtgtccg	gcccggagatc	acaccaagct	gaagatgtac	3180
gcggtacgcc	aaggcaagac	cattaccgag	ctgctatctg	aatacatcgc	gcagctacca	3240
gagtaaatga	gcaaatgaat	aaatgagtag	atgaatttta	gcccgtaaag	gaggcggtcat	3300
ggaaaatcaa	gaacaaccag	gcaccgacgc	cgtggaatgc	cccatgtgtg	gaggaaacggg	3360
cgggttggcca	ggcgtaagcg	gctgggttgt	ctgcccggcc	tgcaatggca	ctggaacccc	3420
caagcccgcg	gaatcgccgt	gacggtcgca	aacctccgg	cccggtaaca	atcgccgcgg	3480
cgctgggtga	tgacctgggt	gagaagttga	aggccgcgca	ggccgcccag	cggaacgcga	3540
tcgaggcaga	agcacgcccc	ggtgaatcgt	ggcaagcggc	cgctgatcga	atccgcгаа	3600
aatcccggca	accgcccggca	gccggtgcgc	cgctcgattag	gaagccgccc	aaggcgacag	3660
agcaaccaga	ttttttcggt	ccgatgctct	atgacgtggg	caccgcgat	agtcgcgagca	3720
tcattggacgt	ggccgttttc	cgtctgtcga	agcgtgaccg	acgagctggc	gaggtgatcc	3780
gctacgagct	tcacagcggg	cacgtagagg	tttcgcagg	gccggccggc	atggccagtg	3840
tgtgggatta	cgacctgggt	ctgatggcgg	tttcccatct	aaccgaatcc	atgaaccgat	3900
accgggaagg	gaagggagac	aagccccggc	gcgtgttccg	tcacacggtt	gcccagctac	3960
tcaagttctg	ccggcgagcc	gatggcgga	agcagaaaga	cgacctggta	gaaacctgca	4020
ttcggttaaa	caccacgcac	gttgccatgc	agcgtacgaa	gaaggccaag	aacggccgcc	4080
tggtgacggg	atccgagggg	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	4140
ccgggcccgc	ggagtacatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	4200
aaggcaagaa	cccgacgctg	ctgacgggtc	accccgatta	ctttttgatc	gatcccggca	4260
tcggccggtt	tctctaccgc	ctggcacgcc	gcgcgcgagg	caaggcagaa	gccagatggt	4320
tggtcaagac	gatctacgaa	cgcaagtggca	gcccgcggaga	gttcaagaag	ttctgtttca	4380
ccgtgcgcaa	gctgatcggg	tcaaatgacc	tgccggagta	cgatttgaag	gaggaggcgg	4440
ggcaggctgg	cccgatccta	gtcatgcgct	accgcaacct	gatcgagggc	gaagcatccg	4500
ccggttccta	atgtacggag	cagatgctag	ggcaaatggc	cctagcaggg	gaaaaagggtc	4560
gaaaagggtct	ctttcctgtg	gatagcacgt	acattgggaa	cccaaagccg	tacattggga	4620
accggaaccc	gtacattggg	aacccaaagg	cgtacattgg	gaaccgggtca	cacatgtaag	4680
tgactgatata	aaaagagaaa	aaaggcgatt	tttccgcta	aaactcttta	aaacttatta	4740
aaactcttaa	aaccgcgctg	gcctgtgcat	aactgtctgg	ccagcgca	gccgaagagc	4800
tgcaaaaagc	gcctaccctt	cggtcgctgc	gctccctacg	ccccgcgct	tcgcgtcgcc	4860
ctatcgccgc	cgctggccgc	tcaaaaatgg	ctggcctacg	gccaggcaat	ctaccagggc	4920
gcccgaacag	cgccgcgctc	ccactcgacc	gccggcgccc	acatcaaggc	acctgcctc	4980
gcccgtttcg	gtgatgacgg	tgaaaacctc	tgaacatgc	agctcccggga	gacggtcaca	5040
gcttgtctgt	aagcggatgc	ccggagcaga	caagcccgctc	aggcgcgctc	agcgggtggt	5100
ggcggtgtc	ggggcgccagc	catgacccag	tcacgtagcg	atagcggagt	gtatactggc	5160
ttaactatgc	ggcatcagag	cagattgtac	tgagagtga	ccatagcgg	tgtgaaatag	5220
cgcacagatg	cgtaaggaga	aaataccgca	tcaggcgctc	ttccgcttcc	tcgctcactg	5280
actcgctgcg	ctcggtcggt	cggctgcggc	gagcgggtatc	agctcactca	aaggcggttaa	5340

tacgggttatc	cacagaatca	ggggataaacg	caggaaagaa	catgtgagca	aaaggccagc	5400
aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggt	tttccatagg	ctccgcccc	5460
ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagagggtg	gcgaaacccg	acaggactat	5520
aaagatacca	ggcggtttccc	cctggaagct	ccctcgtgcg	ctctcctggt	ccgaccctgc	5580
cgcttacccg	atacctgtcc	gcctttctcc	cttcgggaag	cgtggcgctt	tctcatagct	5640
cacgctgtag	gtatctcagt	tccgtgtagg	tcggttcgctc	caagctgggc	tgtgtgcacg	5700
aaccccccg	tcagcccgac	cgtgcgcct	tatccggtaa	ctatcgtctt	gagtccaacc	5760
cggtaaagca	cgacttatcg	ccactggcag	tgctccactgg	taacaggatt	agcagagcga	5820
gggtatgtagg	cggtgctaca	gagttcttga	agtgggtggcc	taactacggc	tacactagaa	5880
ggacagttat	tggtatctgc	gctctgctga	agccagttac	cttcggaaaa	agagttggta	5940
gctcttgatc	cggcaaaaca	accaccgctg	gtagcgggtgg	tttttttggt	tgcaagcagc	6000
agattacgcg	cagaaaaaaa	ggatctcaag	aagatccttt	gatcttttct	acgggggtctg	6060
acgctcagtg	gaacgaaaaa	tcacgttaag	ggattttggt	catgcattct	aggtactaaa	6120
acaattccatc	cagtaaaata	taatatTTTA	ttttctccca	atcagggttg	atccccagta	6180
agtcaaaaaa	tagctcgaca	tactgttctt	ccccgatata	ctccctgatc	gaccggacgc	6240
agaaggcaat	gtcataccac	ttgtccgccc	tgccgcttct	cccaagatca	ataaagccac	6300
ttactttgoc	atctttcaca	aagatgttgc	tgctccccag	gtcgcggtgg	gaaaagacaa	6360
gttcctcttc	gggcttttcc	gtcttttaaa	aatcatacag	ctcgcgcgga	tctttaaatg	6420
gagtgctctt	ttcccagttt	tcgcaatcca	catcggccag	atcgttattc	agtaagtaat	6480
ccaattcggc	taagcggctg	tctaagctat	tcgtataggg	acaatccgat	atgtcgatgg	6540
agtgaagag	cctgatgcac	tccgcataca	gctcgataat	cttttcaggg	ctttgttcat	6600
cttcatactc	ttccgagcaa	aggacgccat	cggcctcact	catgagcaga	ttgtcccagc	6660
catcatgccg	ttcaaagtgc	aggacctttg	gaacaggcag	ctttccttcc	agccatagca	6720
tcattgtcct	ttcccgcttc	acatcatagg	tggtcccttt	ataccggctg	tccgtcattt	6780
ttaaatatag	gttttcatatt	tctcccacca	gcttatatac	cttagcagga	gacattcctt	6840
ccgtatcttt	tacgcagcgg	tatttttcta	tcagtttttt	caattccggt	gatattctca	6900
tttttagccat	ttattatttt	cttccctctt	tctacagtat	ttaaagatac	cccaagaagc	6960
taattataac	aagacgaact	ccaattcact	gttccttgca	ttctaaaacc	ttaaatacca	7020
gaaaacagct	ttttcaaagt	tgttttcaaa	gttggcgctat	aacatagtat	cgacggagcc	7080
gatttttgaaa	cgcgggtgat	cacaggcagc	aacgctctgt	catcgttaca	atcaacatgc	7140
taccctccgc	gagatcatcc	gtgtttcaaa	cccggcagct	tagttgcccgt	tcttccgaat	7200
agcatcggta	acatgagcaa	agtcctgcgc	cttacaacgg	ctctcccgct	gacgccgtcc	7260
cggactgatg	ggctgcctgt	atcgagtggg	gatttttggtc	cgagctgccg	gtcggggagc	7320
tgttggtcgg	ctgggtggcag	gatataattgt	gggtgtaaca	aattgacgct	tagacaactt	7380
aataaacacat	tgccgacggt	tttaattgtac	tgaattaaag	ccgaattaat	tcgggggagc	7440
tggtattttag	tactggattt	tggttttagg	aattagaaat	tttattgata	gaagtatttt	7500
acaaatacaa	atacatacta	aggggtttctt	atatgctcaa	cacatgagcg	aaaccctata	7560
ggaaccctaa	ttcccttatc	tgggaaactac	tcacacatta	ttatggagaa	actcgagctt	7620
gtcgatcgac	agatccgggc	ggcatctact	ctatttcttt	gccctcggac	gagtgtctggg	7680
gcgtcgggtt	ccactatcgg	cgagtacttc	tacacagcca	tcgggtccaga	cggccgcgct	7740
tctgccccgc	atttgtgtac	gcccgcagct	cccggctccg	gatcggacga	ttgcgtcgca	7800
tcgaccttgc	gcccacgctg	catcatcgaa	attgcgctca	accaagctct	gatagagttg	7860
gtcaagacca	atgcggagca	tatacgcccg	gagtcgtggc	gatcctgcaa	gctccggatg	7920
cctccgctcg	aagtagcggg	tctgctgctc	catacaagcc	aaccacggcc	tcagagaaga	7980
gatgtttggc	acctcgattt	gggaatcccc	gaacatcgcc	tcgctccagt	caatgaccgc	8040
tgttatgccc	ccattgtccg	tcaggacatt	gttggagccg	aaatccgctg	gcacgaggtg	8100
ccggacttcg	gggcagtcct	cggcccaaa	catcagctca	tcgagagcct	gcgcgacgga	8160
cgactgacg	gtgtcgctca	tcacagtttg	ccagtgtatc	acatggggat	cagcaatcgc	8220
gcatatgaaa	tcacgccatg	tagtgtattg	accgattcct	tcgggtccga	atgggcccga	8280
cccgtcctgc	tggttaagat	cggccgcgag	gatcgcaccc	atagcctccg	cgaccggttg	8340
tagaacagcg	ggcagttcgg	tttcaggcag	gtcttgcaac	gtgacaccc	gtgcacggcg	8400
ggagatgcaa	taggtcaggc	tctcgctaaa	ctccccaatg	tcaagcactt	ccggaatcgg	8460
gagcgcggcc	gatgcaaagt	gcccataaac	ataacgatct	ttgtagaaac	catcggcgca	8520
gctattttacc	cgcaggacat	atccacgccc	tcctacatcg	aagctgaaag	cacgagattc	8580
ttcgccctcc	gagagctgca	tcaggctcga	gacgctgtcg	aacttttcga	tcagaaactt	8640
ctcgacagac	gtcgcgggtg	gttcaggctt	tttcatatct	cattgcccc	cgggatctgc	8700
gaaagctcga	gagagataga	tttgtagaga	gagactgggtg	atttcagcgt	gtcctctcca	8760
aatgaaatga	acttccttat	atagaggaag	gtcttgcgaa	ggatagtggg	attgtgcgtc	8820
atcccttaag	tcagtgagga	tatcacatca	atccacttgc	tttgaaagcg	tggttggaac	8880
gtctctcttt	tcacagatgc	tcttcgtggg	tgggggtcca	tctttggggac	cactgtcggc	8940
agaggcatct	tgaacgatag	cctttccttt	atcgcaatga	tggcatttgt	aggtgccacc	9000
ttccttttct	actgtccttt	tgatgaagtg	acagatagct	gggcaatgga	atccgaggag	9060
gtttcccgat	attacccttt	gttgaaaagt	ctcaatagcc	ctttggctct	ctgagactgt	9120
atcttttgata	ttcttgaggt	agacgagagt	gtcgtgctcc	accatgttat	cacatcaatc	9180
cacttgcttt	gaagacgtgg	ttggaaacgtc	ttctttttcc	acgatgtctc	tcgtgggtgg	9240
gggtccatct	ttgggaccac	tgccggcagc	ggcatcttga	acgatagcct	ttcctttatc	9300
gcaatgatgg	catttgttagg	tgccaccttc	cttttctact	gtccttttga	tgaagtgaca	9360

gatagctggg	caatggaatc	cgaggaggtt	tcccgatatt	accctttgtt	gaaaagtctc	9420
aatagccctt	tggctcttctg	agactgtatc	tttgatattc	ttggagtaga	cgagagtgtc	9480
gtgctccacc	atgttggcaa	gctgctctag	ccaatacgca	aaccgcctct	ccccgcgcgt	9540
tggccgattc	attaatgcag	ctggcacgac	agggttcccg	actggaaagc	gggcagtggg	9600
cgcaacgcaa	ttaatgtgag	ttagctcact	cattagggac	cccaggcttt	acactttatg	9660
cttccggctc	gtatgtttgtg	tggaaattgtg	agcggataac	aatttcacac	aggaaacagc	9720
tatgaccatg	attacgaatt	cgagctcggg	acccggggat	cctctagagt	cgacctgcag	9780
gcatgcaagc	ttggcactgg	ccgtcgtttt	acaacgtcgt	gactgggaaa	accctggcgt	9840
tacccaactt	aatcgccctt	cagcacatcc	cccttttcgcc	agctggcgta	atagcgaaga	9900
ggcccgccac	gacgcgcctt	cccaacagtt	gcgagcctg	aatggcgaat	gctagagcag	9960
cttgagcttg	gatcagattg	tcgtttcccg	cccttcagttt	agcttcatgg	agtcaaagat	10020
tcaaatagag	gacctaacag	aactcgccgt	aaagactggc	gaacagttca	tacagagtct	10080
cttacgactc	aatgacaaga	agaaaatctt	cgtcaacatg	gtggagcacg	acacacttgt	10140
ctactccaaa	aatatcaaa	atacagcttc	agaagaccaa	agggcaattg	agacttttca	10200
acaaagggtg	atatccggaa	acctcctcgg	attccattgc	ccagctatct	gtcactttat	10260
tgtgaagata	gtggaaaagg	aaggtggctc	ctacaaatgc	catcattgcg	ataaaggaaa	10320
ggccatcggt	gaagatgcct	ctgccgacag	tgggtcccaa	gatggacccc	caccacagag	10380
gagcatcgtg	gaaaaagaag	acgttccaac	cacgtcttca	aagcaagtgg	attgatgtga	10440
tatctccact	gacgtaaggg	atgacgcaca	atcccactat	ccttcgcaag	acccttcctc	10500
tatataagga	agttcatttc	atttggagag	aacacggggg	actcttgac		10549

<210> 93  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> CaMV35SpolyA Primer

<400> 93  
 ctgaattaac gccgaattaa ttccgggggat ctg 33

<210> 94  
 <211> 29  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> CaMV35Spr Primer

<400> 94  
 ctagagcagc ttgccaacat ggtggagca 29

<210> 95  
 <211> 12592  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pAg2 Plasmid

<400> 95						
gtacgaagaa	ggccaagaac	ggccgcctgg	tgacggtatc	cgagggtgaa	gccttgatta	60
gccgtacaa	gatcgtaaa	agcgaaaccg	ggcgcccgga	gtacatcgag	atcgagctag	120
ctgattggat	gtaccgcgag	atcacagaag	gcaagaaccc	ggacgtgctg	acggttcacc	180
ccgattactt	tttgatcgat	cccggcatcg	gccgttttct	ctaccgcctg	gcacgcccg	240
ccgcaaggcaa	ggcagaagcc	agatgggtgt	tcaagacgat	ctacgaacgc	agtggcagcg	300
ccggagagtt	caagaagttc	tgttttcaccg	tgcgcaagct	gatcgggtca	aatgacctgc	360
cggagtacga	tttgaaggag	gaggcggggc	aggctggccc	gacccatagc	atgcgctacc	420
gcaacctgat	cgaggggcgaa	gcacccgcgg	gttcctaatg	tacggagcag	atgctagggc	480
aaattggcct	agcaggggaa	aaaggtcgaa	aaggtctctt	tcctgtggat	agcacgtaca	540
ttgggaaccc	aaagccgtac	attgggaacc	ggaacccgta	catttgggaac	ccaaagccgt	600
acatttgggaa	ccggtcacac	atgtaagtga	ctgatataaa	agagaaaaaa	ggcgattttt	660
ccgcctaaaa	ctctttaaaa	cttattaaaa	ctcttaaaac	ccgcctggcc	tgtgcataac	720
tgtctggcca	gcgcacagcc	gaagagctgc	aaaaagcgcc	tacccttcgg	tcgctgcgct	780
ccctacgccc	cgccgcttcg	cgtcggccta	tcgcggccgc	tggccgctca	aaaatgggtg	840
gcctacggcc	aggcaatcta	ccagggcgcg	gacaagccgc	gccgtcgcca	ctcgaccgcc	900

ggcgcccccaca	tcaaggccacc	ctgcctcgcg	cgtttcgggtg	atgacgggtga	aaacctctga	960
cacatgcagc	tcccggagac	ggtcacagct	tgtctgtaag	cggatgccgg	gagcagacaa	1020
gcccgtcagg	gcccgtcagc	gggtgttgcc	gggtgtcggg	gcgcagccat	gacccagtc	1080
cgtagcgata	gcccgtgta	tactggctta	actatgcggc	atcagagcag	attgtactga	1140
gagtgaccca	tatgcgggtg	gaaataccgc	acagatgcgt	aaggagaaaa	taccgcatca	1200
ggcgctcttc	cgcttcctcg	ctcactgact	cgctgcgctc	ggctcgttcgg	ctgcggcgag	1260
cggtatcagc	tcactcaaag	gcccgtaatc	ggttatccac	agaatcaggg	gataacgcag	1320
gaaagaacat	gtgagcaaaa	ggccagcaaa	agggcaggaa	ccgtaaaaag	gcccgtgtgc	1380
tggcggtttt	ccataggctc	cgcccccttg	acgagcatca	caaaaaatcg	cgctcaagtc	1440
agagggtggc	aaacccgaca	ggactataaa	gataccaggc	gtttccccct	ggaagctccc	1500
tcgtgcgctc	tcctgttccg	accctgccgc	ttaccggata	cctgtccgcc	tttctccctt	1560
cgggaagcgt	ggcgctttct	catagctcac	gctgtaggta	tctcagttcg	gtgtaggctc	1620
ttcgctccaa	gctgggctgt	gtgcacgaac	cccccgttca	gcccagaccg	tgccgcttat	1680
ccggtaacta	tcgtcttgag	tccaacccgg	taagacacga	cttatcgcca	ctggcagcag	1740
ccactggtaa	caggattagc	agagcgagg	atgtaggcgg	tgctacagag	ttcttgaagt	1800
ggtggccctaa	actagcgtac	actagaagga	cagtatgtgg	tatctgcgct	ctgtgaagc	1860
cagttacctt	cggaaaaaga	gttggtagct	cttgatccgg	caaaacaaacc	accgctggta	1920
gcgggtgggtt	ttttgtttgc	aagcagcaga	ttacgcgcag	aaaaaaagga	tctcaagaa	1980
atcctttgat	cttttctacg	gggtctgacg	ctcagtgga	cgaaaactca	cgtaaaggga	2040
ttttgttcgt	gcattctagg	tactaaaaca	attcatccag	taaaaatata	tattttat	2100
tctcccaatc	aggcttgatc	cccagtaagt	caaaaaatag	ctcgacatac	tgttcttccc	2160
cgatatcctc	cctgatcgac	cggacgcaga	aggcaatgtc	ataccacttg	tccgcctgtc	2220
cgttctctcc	aagatcaata	aagccactta	ctttgccatc	tttcacaaag	atgttgctgt	2280
ctcccagggtc	gcccgtgggaa	aagacaagtt	cctcttcggg	cttttccgctc	tttaaaaaat	2340
catcacagctc	gcccggatct	ttaaatggag	tgtcttcttc	ccagttttcg	caatccacat	2400
cggccagatc	gttattcagt	aagtaatcca	attcggctaa	gcccgtgtct	aagctattcg	2460
tatagggaca	atccgatatg	tcgatggagt	gaaagagcct	gatgcactcc	gcatacagct	2520
cgataatctt	ttcagggtct	tgttcatctt	catactcttc	cgagcaaagg	acgccatcgg	2580
cctcactcat	gagcagattg	ctccagccat	catgcggttc	aaagtgcagg	acctttggaa	2640
caggcagctt	tccttccagc	catagcatca	tgtccttttc	ccgttccaca	tcatagggtg	2700
tccctttata	cggcgtgtcc	gtcattttta	aatatagggt	ttcattttct	cccaccagct	2760
tatatacctt	agcaggagac	attccttccg	tatcttttac	gcagcgggat	ttttcgatca	2820
gttttttcaa	ttccgggtgat	attctcattt	tagccattta	ttatttccct	cctcttttct	2880
acagtattta	aagatacccc	aagaagctaa	ttataacaag	acgaactcca	attcactgtt	2940
ccttgcatct	taaaaacctta	aataccagaa	aacagctttt	tcaaagttgt	tttcaaagtt	3000
ggcgtataaac	atagtatcga	cggagccgat	tttgaaaccg	cgggtgatcac	aggcagcaac	3060
gctctgtcat	cgttacaatc	aacatgctac	cctccgcgag	atcatccgtg	tttcaaaacc	3120
ggcagcttag	ttgcccgtct	tcggaatagc	atcggtaaca	tgagcaaagt	ctgcgcctct	3180
acaacggctc	tcccgtcgac	gcccgtcccg	actgatgggc	tgccgtgatc	gagtggtgat	3240
tttgtgccga	gctgcccgtc	ggggagctgt	tggtcggtcg	gtggcaggat	atatgtgtgt	3300
gtaaacaaat	tgacgcttag	acaacttaat	aacacattgc	ggacgttttt	aatgtactga	3360
attaacgccg	aattaattcg	ggggatctgg	atttttagtac	tggtattttg	ttttaggaat	3420
tagaaatttt	attgatagaa	gtattttaca	aatacaataa	cataactaagg	gtttcttata	3480
tgctcaaac	atgagcgaaa	ccctatagga	accctaattc	ccttatctgg	gaactactca	3540
cacattatta	tggagaaaact	cgagtcaaat	ctcgggtgac	ggcaggaccg	gacggggcgg	3600
taccggcagg	ctgaagtcca	gctgccagaa	acccacgtca	tgccagttcc	cgtgcttgaa	3660
gcccggccgc	cgcagcatgc	cgcggggggc	atatccgagc	gcctcgtgca	tgccgacgct	3720
cgggtcggtg	ggcagcccga	tgacagcgac	cacgctcttg	aagccctgtg	cctccaggga	3780
cttcagcagg	tgggtgtaga	gcgtggagcc	cagtcocgtc	cgctgggtgg	ggggggagac	3840
gtacacggtc	gactcggccg	tcagtcgta	ggcgttgctg	gccttccagg	ggcccgcgta	3900
ggcgtatgcc	gocacctcgc	cgtccacctc	ggcgacgagc	cagggatagc	gctcccgcag	3960
acggacgagg	tcgtccgtcc	actcctgcgg	ttcctgcggc	tcggtaacga	agttgaccgt	4020
gcttgtctcg	atgtagtgg	tgacgatgg	gcagaccgcc	ggcatgtccg	cctcgggtgg	4080
acggcggatg	tggcggggc	gtcgttctgg	gctcatggta	gactcgagag	agatagattt	4140
gtagagagag	actgggtgat	tcagcgtgtc	ctctccaaat	gaaatgaact	tccttatata	4200
gaggaaggct	ttgcgaagga	tagtgggatt	gtgcgtcatc	ccttacgtca	gtggagatat	4260
cacatcaatc	cacttgcctt	gaagacgtgg	ttggaacgtc	ttctttttcc	acgatgctcc	4320
tcgtgggtgg	gggtccatct	ttgggaaccac	tgtcggcgaga	ggcatcttga	acgatagcct	4380
ttcctttatc	gcaatgatgg	catttgttag	tgccaccctc	ctttctact	gtccttttga	4440
tgaagtgaca	gatagctggg	caatggaaat	cgaggagggt	tcccgatatt	accctttgtt	4500
gaaaagtctc	aatagccctt	tggtcttctg	agactgtatc	tttgatatcc	ttggagtaga	4560
cgagagtgtc	gtgctccacc	atgtttatcac	atcaatccac	ttgctttgaa	gacgtgggtg	4620
gaacgtcttc	tttttccacg	atgctcctcg	tgggtggggg	tccatctttg	ggaccactgt	4680
cggcagaggc	atcttgaacg	atagcctttc	ctttatcgca	atgatggcat	ttgtagggtg	4740
caccttctct	ttctactgtc	cttttgatga	agtgacagat	agctgggcaa	tggaaatcca	4800
ggaggtttcc	cgatattacc	ctttgttgaa	aagctcaaat	agcccttttg	tcttctgaga	4860
ctgtatcttt	gatattcttg	gagtagacga	gagtgctcgt	ctccaccatg	ttggcaagct	4920

gctctagcca	atcgcgcaaac	cgccctctccc	cgcgcggttg	ccgattcatt	aatgcagctg	4980
gcacgcacagg	tttcccgcact	ggaaagcggg	cagtgcgcgc	aacgcgaatta	atgtgagtta	5040
gctcactcat	taggcaccccc	aggctttaca	ctttatgctt	ccggctcgtg	tggttggtgg	5100
aattgtgagc	ggataacaat	ttcacacagg	aaacagctat	gaccatgatt	acgaattcga	5160
gccttgacta	gaggggtcgac	ggtatacaga	catgataaga	tacattgatg	agtttgagca	5220
aaccacaact	agaatgcagt	gaaaaaaatg	ctttatttgt	gaaatttgtg	atgctattgc	5280
tttatttgtg	accattataa	gctgcaataa	acaagtggg	gtgggcgaag	aactccagca	5340
tgagatcccc	gcgctggagg	atcatccagc	cgcgctcccg	gaaaaacgatt	ccgaagccca	5400
acctttcata	gaaggcggcg	gtggaatcga	aatctcgtag	cacgtgtcag	tcctgctcct	5460
cgccacagaa	gtgcacgcag	ttgccggccg	ggtcgcgag	ggcgaactcc	cgccccacg	5520
gctgctcgcc	gatctcggtc	atggccggcc	cggaagcgtc	ccggaagtcc	gtggacacga	5580
cctccgacca	ctcggcgtac	agctcgtcca	ggccgcgcac	ccacacccag	gccagggtgt	5640
tgtccggcac	cacctggtec	tggaccgcgc	tgtgaacag	ggtcacgtcg	tccccgacca	5700
caccggcgaa	gtcgtctccc	acgaagtccc	ggggaacccc	gagccgggtcg	gtccagaact	5760
cgaccgctcc	ggcgacgtcg	cgcgcggtga	gcaccggaac	ggcactgggtc	aacttggcca	5820
tggatccaga	tttcgctcaa	gttagtataa	aaaagcaggc	ttcaatcctg	caggaattcg	5880
atcgacactc	tcgtctactc	caagaatata	aaagatacag	tctcagaaga	ccaaagggct	5940
attgagactt	ttcaacaaag	ggtaatatcg	ggaaacctcc	tcggattcca	ttgcccagct	6000
atctgtcact	tcatacaaaag	gacagtagaa	aaggaaggtg	gcacctacaa	atgccatcat	6060
tgcgataaag	gaaagggtat	cgcttcaagat	gcctctgccc	acagtgggtcc	caaagattga	6120
ccccacacca	cgaggagcat	cgtggaaaaa	gaagacgttc	caaccacgtc	ttcaaaagcaa	6180
gtggattgat	gtgataaacat	ggtggagcac	gacactctcg	tctactccaa	gaatatcaaa	6240
gatacagctc	cagaagacca	aagggtctatt	gagacttttc	aacaaagggt	aatatcggga	6300
aacctcctcg	gattccattg	cccagctatc	tgtcacttca	tcaaaaggac	agtagaaaag	6360
gaagggtggca	cctacaaaatg	ccatcattgc	gataaaggaa	aggctatcgt	tcaagatgcc	6420
tctgccgaca	gtggtcccaa	agatggaccc	ccaccacga	ggagcatcgt	ggaaaaagaa	6480
gacgttccaa	ccacgtcttc	aaagcaagtg	gattgatgtg	atatctccac	tgacgtaagg	6540
gatgacgcac	aatccacta	tccttcgcaa	gaccttcttc	tatataagga	agttcatttc	6600
attggagag	gacacgctga	aatcaccagt	ctctctctac	aaatctatct	ctctcgagct	6660
ttcgacatc	cggggggggca	atgagatatg	aaaaagcctg	aactcaccgc	gacgtctgtc	6720
gagaagtttc	tgatcgaaaa	gttcgacagc	gtctccgacc	tgatgcagct	ctcggagggc	6780
gaagaatctc	gtgctttcag	cttcgatgta	ggaggcgctg	gatatgtcct	gcgggtaaat	6840
agctgcgcg	atggtttcta	caaagatcgt	tatgtttatc	ggcactttgc	atcgcccgcg	6900
ctcccgatcc	cggaagtgtc	tgacattggg	gagtttagcg	agagcctgac	ctattgcatc	6960
tcccgcgctg	cacagggtgt	cacgttgcaa	gacctgcctg	aaaccgaact	gcccgtgtgt	7020
ctacaacagg	tcgcggaggc	tatggatgag	atcgctgcgg	ccgatcttag	ccagacgagc	7080
gggttcggcc	cattcggacc	gcaaggaaatc	ggtcaatata	ctacatggcg	tgatttcata	7140
tgccgcgctg	ctgatcccca	tgtgtatcac	tgccaaactg	tgatggacga	caccgtcagt	7200
gcgtccgtcg	cgcaggctct	cgatgagctg	atgctttggg	ccgaggactg	ccccgaagtc	7260
cgccacctcg	tgacgcggga	tttcggctcc	aacaatgtcc	tgacggacaa	tgcccgcata	7320
acagcgggtc	ttgactggag	cgaggcgatg	ttcggggatt	cccaatacga	ggtcgccaa	7380
atcttcttct	ggaggccgtg	gttggcttgt	atggagcagc	agacgcgcta	cttcgagcgg	7440
aggcatccgg	agcttgcagg	atcgccacga	ctccggcgct	atatgctccg	cattgggtctt	7500
gaccaactct	atcagagctt	ggttgacggc	aatttcgatg	atgcagcttg	ggcgcagggt	7560
cgatgcgacg	caatcgtccg	atccggagcc	gggactgtcg	ggcgtaacaa	aatcgccgcg	7620
agaagcgcgg	ccgtctggac	cgatggctgt	gtagaagtac	tcgccgatag	tggaaccgga	7680
cgccccagca	ctcgtccgag	ggcaaaagaa	tagagttagat	gccgaccgga	tctgtcgatc	7740
gacaagctcg	agtttctcca	taataatgtg	tgagttagttc	ccagataagg	gaattagggt	7800
tcctataggg	tttcgctcat	gtgttgagca	tataagaaac	ccttagtatg	tatttgtatt	7860
tgtaaaatac	ttctatcaat	aaaatttcta	attcctaaaa	ccaaaatcca	gtactaaaaa	7920
ccagatcccc	cgaattaatt	cggcgttaat	tcagatcaag	cttggcactg	gccgtcgttt	7980
tacaacgtcg	tgactgggaa	aaccctggcg	ttacccaact	taatcgctt	gcagcacatc	8040
cccctttcgc	cagctggcgt	aatagcgaag	aggcccgac	cgatcgccct	tcccaacagt	8100
tgccgagcct	gaatggcgaa	tgctagagca	gcttgagctt	ggatcagatt	gtcgtttccc	8160
gccttcagtt	tggggatcct	ctagactgaa	ggcgggaaac	gacaatctga	tcagtagcgg	8220
agaattaagg	gagtcacggt	atgacccccg	ccgatgcagc	gggacaagcc	gttttacggt	8280
tggaaactgac	agaaccgcaa	cgltgaagga	gccactcagc	cgcgggtttc	tggagtttaa	8340
tgagctaagc	acatacgtca	gaaaccatta	ttgcgcgttc	aaaagtcgcc	taaggctcact	8400
atcagctagc	aaatatctct	tgtcaaaaat	gctccactga	cgttccataa	attccccctg	8460
gtatccaatt	agagtctcat	attcactctc	aatccaaata	atctgcaccg	gatctcgaga	8520
atcgaattcc	cgcggccgcc	atggttagatc	tgactagtaa	aggagaagaa	cttttctactg	8580
gagttgtccc	aattcttgtt	gaattagatg	gtgatgttaa	tgggcacaaa	ttttctgtca	8640
gtggagaggg	tgaaggtgat	gcaacatacg	gaaaacttac	cottaaattt	atttgactaa	8700
ctggaaaact	acctgttccg	tggccaacac	ttgtcactac	tttctcttat	ggtgttcaat	8760
gcttttcaag	ataccagatg	catatgaagc	ggcacgactt	cttcaagagc	gccatgcctg	8820
agggatcgt	gcaggagagg	accatcttct	tcaaggacga	cgggaactac	aagacacgtg	8880
ctgaagtcaa	gtttgagggg	gacaccctcg	tcaacaggat	cgagcttaag	ggaatcgatt	8940

tcaaggagga	cggaaacatc	ctcggccaca	agttggaata	caactacaac	tcccacaacg	9000
tatacatcat	ggccgacaag	caaaaagaacg	gcatcaaagc	caacttcaag	acccgccaca	9060
acatcgaaga	cggcggcgtg	caactcgcgtg	atcattatca	acaaaataact	ccaattggcg	9120
atggccctgt	cctttttacca	gacaaccatt	acctgtccac	acaatctgcc	ctttcgaaaag	9180
atcccaacga	aaagagagac	cacatggtcc	ttcttgagtt	tgtaacagct	gctgggatta	9240
cacatggcat	ggatgaacta	tacaaagcta	gccaccacca	ccaccaccac	gtgtgaattg	9300
gtgaccagct	cgaattttccc	cgatcggtca	aacatttggc	aataaagttt	cttaagattg	9360
aatcctgttg	cgggtcttgc	gatgattatc	atataatttc	tgttgaatta	cgtaagcat	9420
gtaataatta	acatgtaatg	catgacgtta	tttatgagat	gggtttttat	gattagagtc	9480
ccgcaattat	acattttaata	cgcgatagaa	aacaaaatat	agcgcgcaaa	ctaggataaa	9540
ttatcgcgcg	cgggtgtcatc	tatgttacta	gatcgggaat	taaactatca	gtgtttgaca	9600
ggatatattg	gcgggttaaac	ctaagagaaa	agagcggtta	ttagaataac	ggatatattaa	9660
aagggtttat	aaagggtttat	ccgttctgtcc	atttgtatgt	gcatgccaac	cacaggggtc	9720
ccctcgggat	caaaagtactt	tgatccaacc	cctccgctgc	tatagtgcag	tcggcttctg	9780
acgttcctag	cagccgtctt	ctgaaaacga	catgtcgcac	aagtcctaag	ttacgcgaca	9840
gggtccgcgc	ctgccctttt	cctggcggtt	tcttgcgcgc	tggttttagtc	gcataaagta	9900
gaatacttgc	gactagaacc	ggagacatta	cgccatgaac	aagagcgccg	ccgctggcct	9960
gctgggctat	gcccgcgtca	gcaccgacga	ccaggacttg	accaaccaac	gggcccgaact	10020
gcacgcggcc	ggctgcacca	agctgttttc	cgagaagatc	accggcacca	ggcgcgaccg	10080
cccggagctg	gccaggatgc	ttgaccacct	acgcccgtgc	gacgttgtga	cagtgaccag	10140
gctagaccgc	ctggcccgcga	gcacccgcga	cctactggac	attgccgagc	gcatccagga	10200
ggccggcgcg	ggcctgcgtta	gcctggcaga	gccgtgggcc	gacaccacca	cgccggccgg	10260
ccgcatgggt	ttgaccgtgt	tcgcccgcac	tgccgagttc	gagcgttccc	taatcatcga	10320
ccgacccggg	agcgggcgcg	agcgggccaa	ggcccaggcc	gtgaagtttg	gcccccgccc	10380
taccctcacc	ccggcacaga	tcgcccgcac	ccgcccagct	atcgaccagg	aaggccgcac	10440
cgtgaaagag	gcggctgcac	tgcttggcgt	gcacgcctcg	accctgtacc	gcgcacttga	10500
gcgcagcgag	gaagtgcacg	ccaccgagcg	caggcgcgcg	gggtgccttc	gtgaggacgc	10560
attgaccgag	gccgacgccc	tgccggcgcc	cgagaatgaa	cgccaagagg	aacaagcatg	10620
aaaccgcacc	aggacggcca	ggacgaaccc	tttttcatca	ccgaagagat	cgaggcgagg	10680
atgatcgcgg	ccgggtacgt	gttcgagccg	cccgcgcacg	tctcaaccgt	gcggctgcat	10740
gaaatcctgg	ccgggtttgtc	tgatgccaaag	ctggcgccct	ggccggccag	cttggccgct	10800
gaagaaaccg	agcggccgccc	tctaaaaagg	tgatgtgtat	ttgagtaaaa	cagcttgctg	10860
catgcggtcg	ctgcgtatat	gatgcgtatga	gtaaaataaac	aaatacgcga	ggggaaacga	10920
tgaaggttat	cgctgtactt	aaccagaaag	gcgggtcagg	caagacgacc	atcgcaaccc	10980
atctagcccc	cgccctgcaa	ctcgccgggg	ccgatgttct	gttagtcgat	tccgatcccc	11040
agggcagtg	ccgcgatttg	gcggccgtgc	gggaagatca	accgctaacc	gttgtcgga	11100
tcgaccgccc	gacgattgac	cgcgacgtga	aggccatcgg	ccggcgcgac	ttcgtagtga	11160
tcgacggagc	gccccaggcg	gcggacttgg	ctgtgtccgc	gatcaaggca	gccgacttcg	11220
tgctgattcc	gggtgcagcca	agcccttacg	acatatgggc	caccgcgcgc	ctgggtggagc	11280
tggttaagca	gcgcattgag	gtcacggatg	gaaggctaca	agcggccttt	gtcgtgtcgc	11340
ggcgcatcaa	aggcacgcgc	atcggcgggtg	agggtgccga	ggcgctggcc	gggtacgagc	11400
tgccatttcc	tgagtcctcg	atcacgcagc	gcgtgagcta	cccaggcaact	gccgcccgcg	11460
gcacaaccgt	tcttgaatca	gaacccgagg	gcgacgctgc	ccgcgaggtc	caggcgctgg	11520
ccgctgaaat	taaatcaaaa	ctcattttgag	ttaatgaggt	aaagagaaaa	tgagcaaaaag	11580
cacaaacacg	ctaagtgcgc	gcccgtccgag	cgcacgcagc	agcaaggctg	caacgttggc	11640
cagcctggca	gacacgccag	ccatgaagcg	ggtcaacttt	cagttgcccg	cggaggatca	11700
caccaagctg	aagatgtacg	cggtagcgca	aggcaagacc	ataccgagc	tgctatctga	11760
atacatcgcg	cagctaccag	agtaaatgag	caaatgaata	aatgagtaga	tgaattttag	11820
cggctaaagg	aggcggcatg	gaaaatcaag	aacaaccagg	caccgacgcc	gtggaatgcc	11880
ccatgtgtgg	aggaacgggg	ggttggccag	gcgtaagcgg	ctgggttgtc	tgccggccct	11940
gcaatggcac	tggaaccccc	aagcccagg	aatcggcgtg	acggctcgcaa	accatccggc	12000
ccggtacaaa	tcggcgcgcc	gctgggtgat	gacctggtgg	agaagtgaag	ggccgcgcag	12060
gccgcccagc	ggcaacgcac	cgaggcagaa	gcacgcccgc	gtgaatcgtg	gcaagcggcc	12120
gctgatcgaa	tcggcaaaaga	atcccggcaa	ccgcccgcag	ccggtgcgcc	gtcgattagg	12180
aagccgcccc	agggcgacga	gcaaccagat	tttttctgtc	cgatgctcta	tgacgtgggc	12240
acccgcgata	gtcgcagcat	catggacgtg	gccgttttcc	gtctgtcgaa	gcgtgaccga	12300
cgagctggcg	aggtgatccg	ctacgagctt	ccgacgggc	acgtagaggt	ttccgcaggg	12360
ccggccggca	tgccagtggt	gtgggattac	gacctggtac	tgatggcggt	ttcccatcta	12420
accgaatcca	tgaaccgata	ccgggaaggg	aaggagagaa	agcccgcccg	cggttccgt	12480
ccacacgttg	cggacgtact	caagttctgc	cggcgagccg	atggcggaag	gcagaaagac	12540
gacctggtag	aaacctgcat	tcgggttaaac	accacgcacg	ttgccatgca	gc	12592

<210> 96  
 <211> 3357  
 <212> DNA  
 <213> Artificial Sequence



<220>

<223> pGEMEasyNOS Plasmid

<400> 96

```
tatcactagt gaattcgcg cgcctgcag gtcgaccata tgggagagct cccaacgcgt 60
tggatgcata gcttgagtat tctatagtg cactaaata gcttggcgta atcatgggtca 120
tagctgtttc ctgtgtgaaa ttgttatccg ctcaaatc cacaacat acgagccgga 180
agcataaagt gtaaagcctg ggggtgcctaa tgagttagct aactcacatt aattgcgttg 240
cgctcactgc cgcgtttcca gtcgggaaac ctgtcgtgcc agctgcatta atgaatcggc 300
caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt ccgcttcctc gctcactgac 360
tcgctgcgct cggtcgttcg gctgcggcga gcggtatcag ctactcaaa ggcggtataa 420
cggttatcca cagaatcagg ggataacgca ggaaagaaca tgtgagcaaa aggccagcaa 480
aaggccagga accgtaaaaa ggccgcgttg ctggcgcttt tccataggct ccgccccct 540
gacgagcatc acaaaaatcg acgctcaagt cagaggtggc gaaacccgac aggactataa 600
agataccagg cgtttccccc tggaaagctcc ctcgctgcgt ctctgttcc gaccctgccg 660
cttaccggat acctgtccgc ctttctccct tcgggaagcg tggcgctttc tcatagctca 720
cgctgtaggc atctcagttc ggtgtagggt gctcgtcca agctgggctg tgtgcacgaa 780
cccccgcttc agcccgaccg ctgcgcctta tccggtaact atcgcttga gtccaacccg 840
gtaagacagc acttatcgcc actggcagca gccactggta acaggattag cagagcgagg 900
gatgtaggcg gtgctacaga gttcttgaa gttctggccta actacggcta cactagaaga 960
acagtatttg gtatctgcgc tctgtgaag ccagttacct tcggaaaaag agttggtagc 1020
tcttgatccg gcaaacaaac caccgctggg agcggtgggt tttttgtttg caagcagcag 1080
attacgcgca gaaaaaaagg atctcaagaa gatcctttga tctttctac ggggtctgac 1140
gctcagtgga acgaaaaact acgttaaggg attttggtca tgagattatc aaaaaggatc 1200
ttcacctaga tctttttaa ttaaaaaatga agttttaaat caatctaaag tatatatgag 1260
taaaacttgg ctgacagtta ccaatgctta atcagtgagg cactatctc agcgatctgt 1320
ctatttcggt catccatagt tgctgactc cccgtcgtgt agataactac gatacgggag 1380
ggcttaccat ctggcccccag tgctgcaatg ataccgcgag acccacgctc accggctcca 1440
gatttatcag caataaacca gccagccgga agggccgagc gcagaagtgg tcttgcaact 1500
ttatccgcct ccacccagtc tattaattgt tgccgggaag ctagagtaag tagttcgcca 1560
gttaatatgt tgcgcaacgt tggtgccatt gctacaggca tcgtgggtgc acgctcgtcg 1620
tttggtatgg cttcattcag ctccggttcc caacgataca ggcgagttac atgatcccc 1680
atgttgtgca aaaaagcggg tagctccttc ggtcctccga tcgttgtcag aagtaagttg 1740
gcccgcatgt tatcactcat ggttatggca gctactgata attctcttac tgtcatgcca 1800
tccgtaagat gcttttctgt gactgggtgag tactcaacca agtcattctg agaatagtgt 1860
atgcccgcac cgagttgctc ttgcccggcg tcaatacggg ataataccgc gccacatagc 1920
agaactttaa aagtgtctat cattggaaaa cgttcttcgg ggcgaaaaact ctcaaggatc 1980
ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg caccacaactg atcttcagca 2040
tcttttactt tcaccagcgt ttctgggtga gcaaaaaacag gaaggcaaaa tgccgcaaaa 2100
aagggaataa gggcgacacg gaaatgttga atactcatac tcttctttt tcaatattat 2160
tgaagcattt atcagggtta ttgtctcatg agcggataca tatttgaatg tatttagaaa 2220
aataaaca aa taggggttcc gcgcacattt ccccgaaaag tgccacctga tgcggtgtga 2280
aataccgcac agatgcgtaa ggagaaaaata ccgcatcagg aaattgtgaa cgtaatatatt 2340
ttgttaaaat tcgcgttaaa tttttgttaa atcagctcat tttttaacca ataggccgaa 2400
atcggcaaaa tcccttataa atcaaaaagaa tagaccgaga taggggttgag tgttgttcca 2460
gtttggaaca agagtccact attaaagaac gtggactcca acgtcaaagg gcgaaaaacc 2520
gtctatcagg gcgatggccc actacgtgaa ccatcaccct aatcaagttt ttgggggtcg 2580
aggtgccgta aagcactaaa tcggaaccc taaagggagcc cccgatttag agcttgacgg 2640
ggaaagccgg cgaacgtggc gagaaaaggaa gggaaagaa cgaagaggagc gggcgctagg 2700
gcgctggcaa gtgtagcggg cacgctgcgc gtaaccacca caccgcgcg gcttaatgcg 2760
ccgctacagg cgcgctccat tcgccattca ggctgcgcaa ctgttgggaa gggcgatcgg 2820
tgccggcctc ttcgctatta cgcagctgg cgaaaggggg atgtgctgca aggcgattaa 2880
gttgggtaac gccaggggtt tcccagtcac gacgttgtaa aacgacggcc agtgaattgt 2940
aatagactc actatagggc gaattgggccc cgacgtcgca tgctcccggc cgccatggcg 3000
gccgcgggaa ttcgattctc gagatccggg gcagattatt tggattgaga gtgaatatga 3060
gactctaatt ggataccgag gggaaatttat ggaacgtcag tggagcattt ttgacaagaa 3120
atatttgcta gctgatagt accttaggcg acttttgaac gcgcaataat ggtttctgac 3180
gtatgtgctt agctcattaa actccagaaa cccgcggctg agtggtcctc tcaacgttgc 3240
ggttctgtca gttccaaacg taaaacggct tgtcccgcgt catcgccggg ggtcataacg 3300
tgactccctc aattctccgc tcatgatcag attgtcggtt cccgccttca gtctaga 3357
```

<210> 97

<211> 10122

<212> DNA

<213> Artificial Sequence

<220>



<223> p1302NOS Plasmid

<400> 97

catggtatgat	ctgactagta	aaggagaaga	actttttcact	ggagttgtcc	caattcttgt	60
tgaattagat	ggtgatgtta	atgggcacaa	atttttctgtc	agtggagagg	gtgaaggtga	120
tgcaacatac	ggaaaactta	cccttaaatt	tattttgcact	actggaaaac	tacctgttcc	180
gtggccaaca	cttgtcacta	ctttctctta	tggtgttcaa	tgcttttcaa	gataccaga	240
tcatatgaag	cggcacgact	tcttcaagag	cgccatgcct	gagggatagc	tgaggagag	300
gaccatcttc	ttcaaggacg	acgggaacta	caagacacgt	gctgaagtca	agtttgagg	360
agacaccctc	gtcaacagga	tcgagcttaa	gggaatcgat	ttcaaggagg	acggaaacat	420
cctcggccac	aagttggaat	acaactacaa	ctcccacaac	gtatacatca	tgcccgacaa	480
gcaaaagaac	ggcatcaaa	ccaacttcaa	gacccgccac	aacatcgaag	acggcggcgt	540
gcaactcgct	gatcattatc	aacaaaatac	tccaattggc	gatggccctg	tcctttttacc	600
agacaaccat	tacctgtcca	cacaatctgc	cctttcgaaa	gatcccaacg	aaaagagaga	660
ccacatggtc	cttcttgagt	ttgtaacagc	tgctgggatt	acacatggca	tggtatgaat	720
atacaaaagt	agccaccacc	accaccacca	cgtgtgaatt	ggtgaccagc	tcgaatttcc	780
ccgatcggtc	aaacatttgg	caataaagtt	cctttaaagatt	gaatcctgtt	gcgggtcttg	840
cgatgattat	catataaatt	ctggtgaatt	acgttaagca	tgtaataatt	aacatgtaat	900
gcatgacgtt	atttatgaga	tgggttttta	tgatttagagt	cccgaatta	tacatttaat	960
acgcatgaga	aaacaaaata	tagcgcgcaa	ctaggataaa	attatcgcgc	gcggtgtcat	1020
ctatgttact	agatcgggaa	ttaaaactatc	agtgtttgac	aggatatatt	ggcgggtaaa	1080
cctaagagaa	aagagcggtt	attagaataa	cggatattta	aaagggcggt	aaaaggttta	1140
tcogttcgtc	caattgtatg	tgcatgccaa	ccacagggtt	cccctcggga	tcaaagtact	1200
ttgatccaac	ccctccgctg	ctatagtgc	gtcggcttct	gacgttcagt	gcagccgtct	1260
tctgaaaacg	acatgtcgca	caagtccata	gttacgcgac	aggctgccc	cctgcccctt	1320
tcctggcggt	ttcttgtcgc	gtgttttagt	cgcataaaagt	agaataactt	cgactagaac	1380
cggagacatt	acgccatgaa	caagagcgcc	gccgctggcc	tgctgggcta	tgcccgcgtc	1440
agcaccgagc	accaggactt	gaccaaccac	cgggcccgaac	tgacgcggcg	cggctgcacc	1500
aagctgtttt	ccgagaagat	caccggcacc	agggcgagcc	gcccggagct	ggccaggatg	1560
cttgaccacc	tacgccctgg	cgacgttgtg	acagtgacca	ggctagaccg	cctggcccgc	1620
agcaccgcgc	acctaactga	caattgcccag	cgcacccagg	aggccggcgc	gggcccgtgc	1680
agcctggcag	agccgtgggc	cgacaccacc	acgcgggccc	gcccgcattg	gttgaccgtg	1740
ttcgccggca	ttgccgagtt	cgagcgttcc	ctaactcatc	accgcaccgc	gagcggggcg	1800
gagggccgca	aggcccagag	cgtgaagttt	ggcccccgcc	ctaccctcac	cccggcacag	1860
atcgccgacg	cccgcgagct	gatcgaccag	gaaggccgca	ccgtgaaaag	ggcggtgacg	1920
ctgcttggcg	tgcatcgctc	gaccctgtac	cgcgcacttg	agcgcagcga	ggaagtgcag	1980
cccaccgagg	ccaggcggcg	cgggtgcctt	cgtgaggacg	cattgaccga	ggccgacgcc	2040
ctggcgggcc	ccgagaatga	acgccaagag	gaacaagcat	gaaaccgcac	caggacggcc	2100
aggacgaacc	gtttttcatt	accgaagaga	tcgagggcga	gatgatcgcg	gccgggtacg	2160
tgctcgagcc	gcccgcgcac	gtctcaaccg	tgccgctgca	tgaatacctg	gccgggttgt	2220
ctgatgccaa	gctggcggcc	tggccggcca	gcttggccgc	tgaagaaacc	gagcgcggcc	2280
gtctaaaaag	gtgatgtgta	tttgagttaa	acagcttgcg	tcacgcggtc	gctgcgtata	2340
tgatgcgatg	agtaataaaa	caaatacgca	aggggaacgc	atgaagggtta	tcgctgtact	2400
taaccagaaa	ggcgggtcag	gcaagacgac	catcgcaacc	catctagccc	gcgcccctga	2460
actcgccggg	gccgatgttc	tggttagtca	ttccgatccc	cagggcagtg	cccgcgattg	2520
ggcggccggt	cgggaagatc	aaccgctaac	cgttgtcgcc	atcgaccgcc	cgacgattga	2580
ccgcgacgtg	aaggccatcg	gccggcgcga	cttcgtagtg	atcgacggag	cgccccaggc	2640
ggcggacttg	gctgtgtccg	cgatcaaggc	agccgacttc	gtgctgattc	cgggtgcagc	2700
aagcccttac	gacatatggg	ccaccgcccga	cctggtggag	ctggttaagc	agcgcattga	2760
ggtcacggat	ggaaggctac	aagcggcctt	tgtcgtgtcg	cgggcatgca	aaggcacgcg	2820
catcgccggt	gaggttgccg	aggcgctggc	cgggtacgag	ctgcccattc	ttgagtcgcc	2880
tatcacgcag	cgcgtgagct	accaggcac	tgccgcggcc	ggcacaaccg	ttcttgaatc	2940
agaacccgag	ggcgacgctg	cccgcgaggt	ccaggcgctg	gccgctgaaa	ttaaatcaaa	3000
actcatttga	gttaatgagg	taaagagaaa	atgagcaaaa	gcacaaacac	gctaagtgc	3060
ggcgcgtccg	gcgcacgcag	cagcaaggct	gcaacgttgg	ccagcctggc	agacacgcca	3120
gccaatgaag	gggtcaactt	tcagtgtccg	gcggaggatc	acaccaagct	gaagatgtac	3180
gcggtagccc	aaggcaagac	cattaccgag	ctgctatctg	aatacatcgc	gcagctacca	3240
gagtaaatga	gcaaatgaat	aaatgagtag	atgaatttta	gcggctaaag	gagggcgcat	3300
ggaaaatcaa	gaacaaccag	gcaccgacgc	cgtggaatgc	cccatgtgtg	gaggaacggg	3360
cggttggcca	ggcggttagc	gctgggttgt	ctgcccggcc	tgcaatggca	ctggaacccc	3420
caagcccag	gaatcgcgct	gacggctcgca	aaccatccgg	cccgttaca	atcggcgcgg	3480
cgctgggtga	tgacctgggt	gagaagttga	agggccgcga	ggccgcccag	cggcaacgca	3540
tcgaggcgca	agcaacgccc	ggtgaatcgt	ggcaagcggc	cgctgatcga	atccgcaag	3600
aatcccggca	accgcccggca	gccgggtgcgc	cgtcgattag	gaagccgccc	aaggcgacg	3660
agcaaccaga	ttttttcgtt	ccgatgctct	atgacgtggg	caccccgcat	agtcgcagca	3720
tcatggacgt	ggccggtttt	cgtctgtcga	acgctgaccg	acgagctggc	gaggtgatcc	3780
gctacgagct	tccagacggg	cacgtagagg	tttccgcagg	gccggccggc	atggccagtg	3840

tgtgggatta	cgacctggta	ctgatggcgg	tttcccatct	aaccgaatcc	atgaaccgat	3900
accgggaagg	gaaggagagac	aagcccgggc	gcgtgtttccg	tccacacgtt	gcggaacgtac	3960
tcaagttctg	ccggcgagcc	gatggcggaa	agcagaaaga	cgacctggta	gaaacctgca	4020
ttcgggttaa	caccacgcac	gttgccatgc	agcgtacgaa	gaaggccaag	aacggccgcc	4080
tgggtgacgg	atccgagggg	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	4140
ccgggagggc	ggagtagatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	4200
aaggcaagaa	ccgggacgtg	ctgacggttc	accccgatta	ctttttgatc	gatcccgga	4260
tccggcgttt	tctctaccgc	ctggcacggc	gcggccgagg	caaggcagaa	gccagatggg	4320
tgttcaagac	gatctacgaa	cgagtgaggc	gcggcggaga	gttcaagaag	ttctgtttca	4380
ccgtgcgcaa	gctgatcggg	tcaaatgacc	tgccggagta	cgatttgaag	gaggaggcgg	4440
ggcaggctgg	cccgatccta	gtcatgcgct	accgcaacct	gatcgagggc	gaagcatccg	4500
ccgggttcta	atgtacggag	cagatgctag	ggcaaatgtc	cctagcaggg	gaaaaagggtc	4560
gaaaagggtc	ctttcctgtg	gatagcacgt	acattgggaa	cccaaagccg	tacattggga	4620
accggaaccc	gtacattggg	aacccaagc	cgtagattgg	gaaccgggtc	cacatgttaag	4680
tgactgatat	aaaagagaaa	aaaggcgatt	tttccgccta	aaactcttta	aaacttatta	4740
aaactcttaa	aacccgcctg	gcctgtgcac	aactgtctgg	ccagcgcaca	gccgaagagc	4800
tgcaaaaagc	gcctaccctt	cggtcgctgc	gctccctacg	ccccgccgct	tcgcgtcgcc	4860
ctatcgcgcc	cgctggccgc	tcaaaaatgg	ctggccctacg	gccaggcaat	ctaccagggc	4920
gcggaacaag	cgcgccgctg	ccactcgacc	gcgggcggcc	acatcaaggc	accctgcctc	4980
gcgcgtttcg	gtgatgacgg	tgaaaacctc	tgacacatgc	agctcccggg	gacgggtaca	5040
gcttgtctgt	aagcggatgc	cgggagcaga	caagcccgtc	agggcgcgct	agcgggtgtt	5100
ggcgggtgtc	ggggcgcgag	catgacccag	tcacgtagcg	atagcggagt	gtatactggc	5160
ttactatgc	ggcatcagag	cagattgtac	tgagagtgca	ccatatgcgg	tgtgaaatac	5220
cgacacagatg	cgtaaggaga	aaataccgca	tcaggcgctc	ttccgcttcc	tcgctcactg	5280
actcgctgcg	ctcggtcggt	cggtcgcgcc	gagcgggtatc	agctcactca	aaggcggtaa	5340
tacgggttatc	cacagaatca	ggggataacg	caggaaaagaa	catgtgagca	aaaggccagc	5400
aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggt	tttccatagg	ctccgcccc	5460
ctgacgagca	tcaaaaaaat	cgacgctcaa	gtcagaggtg	gcgaaacccg	acaggactat	5520
aaagatacca	ggcggtttccc	cctggaagct	ccctcgctgc	ctctcctggt	ccgacctgc	5580
cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	cggtggcgctt	tctcatagct	5640
cacgctgtat	gtatctcagt	tcggtgtagg	tcgttcgctc	caagctgggc	tgtgtgcacg	5700
aacccccgt	tcagcccgac	cgctgcgctc	tatccggtaa	ctatcgtctt	gagtcacaac	5760
cggtaaagca	cgacttatcg	ccactggcag	cagccactgg	taacaggatt	agcagagcga	5820
ggtagtagg	cggtgctaca	gagttcttga	agtgggtggc	taactacggc	tacactagaa	5880
ggcagttatc	tggtagtctg	gctctgttga	agccagttac	cttcggaaaa	agagttggt	5940
gctcttgatc	cggaacacaa	accaccgctg	gtagcgggtg	tttttttgtt	tgcaagcagc	6000
agattacgcg	cagaaaaaaa	ggatctcaag	aagatccttt	gatcttttct	acggggtctg	6060
acgctcagtg	gaacgaaaac	tcacgtttaa	ggatttttgt	catgcattct	aggtacctgc	6120
acaattcctc	cagtaaaata	taatatttta	ttttctccca	atcaggcttg	atccccagta	6180
agtcaaaaaa	tagctcgaca	tactgttctt	ccccgatatc	ctccctgatc	gaccggagcg	6240
agaaggcaat	gtcataccac	ttgtccgccc	tgcccttctt	cccaagatca	ataaagcaac	6300
ttactttggc	atctttcaca	aagatgttgc	tgctctcccg	gtcgccgtgg	gaaaagacaa	6360
ggttctcttc	gggcttttcc	gtctttaaia	aatcatacag	ctcgcgcgga	tctttaaatg	6420
gagtgtcttc	ttcccagttt	tcgcaatcca	catcgccag	atcgttattc	agtaagtaat	6480
ccaattccgc	taagcgctg	tctaagcgtat	ctgtatagg	acaatccgat	atgtcgatgg	6540
agtgaagag	cctgatgcac	tcgcgatata	gctcgataat	cttttcagg	ctttgttcat	6600
cttcatactc	ttccgagcaa	aggacgccat	cgccctcact	catgagcaga	ttgctccagc	6660
catcatgcgc	ttcaaagtgc	aggaccttgg	gaacaggcag	ctttccttcc	agccatagca	6720
tcattgtctt	ttcccgttcc	acatcatagg	tggtcccttt	ataccggctg	tccgtcattt	6780
ttaaatatag	gttttcattt	tctcccacca	gcttatatac	cttagcagga	gacattcctt	6840
ccgtatcttt	tacgcagcgg	tatttttcga	tcagtttttt	caattccggg	gatattctca	6900
tttttagccat	ttattatttc	cttccctctt	tctacagtat	ttaaagatac	cccaagaagc	6960
taattataac	aagacgaact	ccaattcact	gttcccttga	ttctaaaacc	ttaaatacca	7020
gaaaacagct	ttttcaaagt	tgttttcaaa	gttggcggtat	aacatagtat	cgacggagcc	7080
gattttgaaa	ccgcgggtgat	cacaggcagc	aacgctctgt	catcgttaca	atcaacatgc	7140
taccctccgc	gagatcatcc	gtgtttcaaa	cccgccagct	tagttgccgt	tcttccgaat	7200
agcatcggt	acatgagcaa	agtctgccc	cttacaacgg	ctctcccgt	gacgcccgtc	7260
cgactgatg	ggctgcctgt	atcgagtgg	gattttgtgc	cgagctgccg	gtcggggagc	7320
tgttggtgg	ctggtggcag	gatataattg	gggtgtaaa	aattgacgct	tagacaactt	7380
aataacacat	tcgggagctt	tttaattgac	tgaaattaac	ccgaattaat	tcgggggagc	7440
tggatttttag	tactggattt	tggtttttag	aattagaaat	tttattgata	gaagtatttt	7500
acaaatacaa	atacatacta	agggtttctt	atatgctcaa	cacatgagcg	aaaccctata	7560
ggaacctaat	ttcccttata	tggaacttac	tcacacatta	ttatggagaa	actcgagctt	7620
gtcgatcgac	agatccgggt	ggcatctact	ctatttcttt	gccctcggac	gagtgcctgg	7680
gcgtcggtt	ccactatcgg	cgagtacttc	tacacagcca	tcggtccaga	cggccgcgct	7740
tctcgggcg	atttgtgtac	gcccagacgt	cccggctccg	gatcggaaga	ttgcgtcgca	7800
tcgacctgc	gcccagctg	catcatcgaa	attgcccgtc	accaagctct	gatagagttg	7860

```

gtcaagacca atgctggagca tatacgcccc gagtcgtggc gatcctgcaa gctccggatg 7920
cctccgctcg aagtagcgcg tctgctgctc catacaagcc aaccacggcc tccagaagaa 7980
gatgtttggcg acctcgctatt gggaatcccc gaacatcgcc tcgctccagt caatgaccgc 8040
tgttatgctgg ccattgtccg tcaggacatt gttggagccg aaatccgctg gcacgaggtg 8100
ccggacttcg gggcagtcct cggcccaaaag catcagctca tcgagagcct gcgcgacgga 8160
cgactgacg gtgtcgctcca tcacagtttg ccagtgatac acatggggat cagcaatcgc 8220
gcatatgaaa tcacgccatg tagtgtattg accgattcct tgcgggtccga atgggccgaa 8280
cccgctcgctc tggctaagat cggccgcagc gatcgcctcc atagcctccg cgaccgggtg 8340
tagaacagcg ggcagttcgg tttcaggcag gtcttgcaac gtgacaccct gtgcacggcg 8400
ggagatgcaa taggtcaggc tctcgctaaa ctcccaatg tcaagcactt ccggaatcgg 8460
gagcgcgggcc gatgcaaagt gccgataaac ataacgatct ttgtagaaac catcggcgca 8520
gctattttacc cgcaggacat atccacgccc tcctacatcg aagctgaaaag cacgagattc 8580
ttcgccctcc gagagctgca tcaggtcggg gacgtgtcg aacttttcga tcagaaactt 8640
ctcgacagac gtgcgggtga gttcaggctt ttccatctct cattgcccc ccggactctg 8700
gaaagctcga gagagataga tttgtagaga gagactgggtg atttcagcgt gtccctctcca 8760
aatgaaatga acttccttat atagaggaag gtcttgcgaa ggatagtggg attgtgcgtc 8820
atcccttacg tcagtggaga tatcacatca atccacttgc tttgaagacg tggttggaac 8880
gtcttctctt tccacgatgc tcctcgctggg tgggggtcca tctttgggac cactgtcggc 8940
agaggcatct tgaacgatag cctttccttt atcgcaatga tggcatttgt aggtgccacc 9000
ttcctttctt actgtccttt tgatgaagtg gggcaatgga atccgaggag 9060
gtttcccgat attacccttt gttgaaaagt ctcaatagcc ctttgggtct ctgagactgt 9120
atcttttgata ttcttgaggat agacgagagt gtcgtgctcc accatgttat cacatcaatc 9180
cacttgcttt gaagacgtgg ttggaacgtc ttctttttcc acgatgctcc tcgtgggtgg 9240
gggtccatct ttgggaccac tgctcggcaga ggcactctga acgatagcct ttcttttatc 9300
gcaatgatgg catttgtagg tgccaccttc cttttctact gtccttttga tgaagtgaca 9360
gatagctggg caatggaatc cgaggaggtt tcccgatatt accctttgtt gaaaagtctc 9420
aatagccctt tggctctctg agactgtatc tttgatattc ttggagtaga cgagagtgtc 9480
gtgtccacc atgttggcaa gctgctctag ccaatacgca aaccgcctct ccccgcgctg 9540
tggccgattc attaatgcag ctggcacgac aggtttcccg actggaaagc gggcgaggag 9600
cgcaacgcaa ttaatgtgag ttagctcact cattaaggac cccaggtctt acactttatg 9660
cttccggctc gtatgttgtg tgggaattgt agcggaataa aatttcacac aggaaacagc 9720
tatgaccatg attacgaatt cgagctcgtt acccggggat cctctagact gaaggcggga 9780
aacgacaatc tgatcatgag cggagaatta agggagtcac gttatgaccc cgcgccgtga 9840
cgccgggacaa gccgttttac gtttggaaat gacagaaccg caacgttgaa ggagccactc 9900
agccgcgggt ttctggagtt taatgagcta agcacatacg tcagaaacca ttattgcgcg 9960
ttcaaaagtc gcctaaggct actatcagct agcaaataat tcttgtcaaa aatgctccac 10020
tgacgttcca taaattcccc tcggtatcca attagagtct catattcact ctcaatccaa 10080
ataatctgca ccggatctcg agaatcgaat tcccgcggcc gc 10122

```

<210> 98  
 <211> 621  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> N. tabacum rDNA intergenic spacer (IGS) sequence

<300>  
 <308> Genbank #Y08422  
 <309> 1997-10-31

```

<400> 98
gtgctagcca atgtttaaca agatgtcaag cacaatgaat gttggtggtt ggtggtcgtg 60
gctggcggtg gtggaaaatt gccgtggttc gagcggtagt gatcggcgat ggttggtgtt 120
tgcagcggtg tttgatatcg gaatcactta tgggtggttg cacaatggag gtgcgtcatg 180
gttattggtg gttggtcatc tatatatttt tataataata ttaagtattt tacctatttt 240
ttacatattt tttattaaat ttatgcattg tttgtatttt taaatagttt ttatcgtaact 300
tggtttataa aatattttat tattttatgt gttatattat tacttgatgt attggaaatt 360
ttctccattg ttttttctat attttataata attttcttat ttttttttgt tttattatgt 420
attttttcgt tttataataa atattttatta aaaaaaatat tatttttgta aaatatatca 480
tttacaatgt ttaaaagtca tttgtgaata tattagctaa gttgtacttc tttttgtgca 540
tttggtgttg tacatgtcta ttatgattct ctggccaaaa catgtctact cctgtcactt 600
gggttttttt ttttaagaca t

```

<210> 99  
 <211> 25  
 <212> DNA

<213> Artificial Sequence

<220>

<223> NTIGS-F1 Primer

<400> 99

gtgctagcca atgtttaaca agatg

25

<210> 100

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> NTIGS-R1 Primer

<400> 100

atgtcttaaa aaaaaaaccc caagtgc

28

<210> 101

<211> 233

<212> DNA

<213> Mus Musculus

<300>

<308> Genbank #V00846

<309> 1989-07-06

<400> 101

gacctggaat atggcgagaa aactgaaaat cacggaaaat gagaaataca cacttttagga 60  
cgtgaaatat ggcgaggaaa actgaaaaag gtggaaaatt tagaaatgtc cactgtagga 120  
cgtggaatat ggcaagaaaa ctgaaaatca tggaaaatga gaaacatcca cttgacgact 180  
tgaaaaatga cgaaatcact aaaaaacgtg aaaaatgaga aatgcacact gaa 233

<210> 102

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> MSAT-F1 Primer

<400> 102

aataccgcgg aagcttgacc tggaatatcg c

31

<210> 103

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> MSAT-R1 Primer

<400> 103

ataaccgcgg agtccttcag tgtgcat

27

<210> 104

<211> 277

<212> DNA

<213> Artificial Sequence

<220>

<223> Nopaline Synthase Promoter Sequence

<300>

<308> Genbank #U09365

<309> 1997-10-17

```

<400> 104
gagctcgaat ttccccgata gttcaaacaat ttggcaataa agttttcttaa gattgaatcc 60
tggtgccggg cttgcgatga ttatcatata atttctgttg aattacgtta agcatgtaat 120
aattaacatg taatgcatga cggtatttat gagatgggtt tttatgatta ggtccccgca 180
attatacatt taatacgcga tagaaaacaa aatatagcgc gcaaactagg ataaattatc 240
gcgcgcgggtg tcattctatgt tactagatcg ggaattc 277

```

```

<210> 105
<211> 1812
<212> DNA
<213> Escherichia coli

```

```

<220>
<221> CDS
<222> (1)...(1812)
<223> Beta-Glucuronidase

```

```

<300>
<308> Genbank #S69414
<309> 1994-09-23

```

```

<400> 105
atg tta cgt cct gta gaa acc cca acc cgt gaa atc aaa aaa ctc gac 48
Met Leu Arg Pro Val Glu Thr Pro Thr Arg Glu Ile Lys Lys Leu Asp
1 5 10 15

ggc ctg tgg gca ttc agt ctg gat cgc gaa aac tgt gga att gat cag 96
Gly Leu Trp Ala Phe Ser Leu Asp Arg Glu Asn Cys Gly Ile Asp Gln
20 25 30

cgt tgg tgg gaa agc gcg tta caa gaa agc cgg gca att gct gtg cca 144
Arg Trp Trp Glu Ser Ala Leu Gln Glu Ser Arg Ala Ile Ala Val Pro
35 40 45

ggc agt ttt aac gat cag ttc gcc gat gca gat att cgt aat tat gcg 192
Gly Ser Phe Asn Asp Gln Phe Ala Asp Ala Asp Ile Arg Asn Tyr Ala
50 55 60

ggc aac gtc tgg tat cag cgc gaa gtc ttt ata ccg aaa ggt tgg gca 240
Gly Asn Val Trp Tyr Gln Arg Glu Val Phe Ile Pro Lys Gly Trp Ala
65 70 75 80

ggc cag cgt atc gtg ctg cgt ttc gat gcg gtc act cat tac ggc aaa 288
Gly Gln Arg Ile Val Leu Arg Phe Asp Ala Val Thr His Tyr Gly Lys
85 90 95

gtg tgg gtc aat aat cag gaa gtg atg gag cat cag ggc ggc tat acg 336
Val Trp Val Asn Asn Gln Glu Val Met Glu His Gln Gly Gly Tyr Thr
100 105 110

cca ttt gaa gcc gat gtc acg ccg tat gtt att gcc ggg aaa agt gta 384
Pro Phe Glu Ala Asp Val Thr Pro Tyr Val Ile Ala Gly Lys Ser Val
115 120 125

cgt atc acc gtt tgt gtg aac aac gaa ctg aac tgg cag act atc ccg 432
Arg Ile Thr Val Cys Val Asn Asn Glu Leu Asn Trp Gln Thr Ile Pro
130 135 140

ccg gga atg gtg att acc gac gaa aac ggc aag aaa aag cag tct tac 480
Pro Gly Met Val Ile Thr Asp Glu Asn Gly Lys Lys Lys Gln Ser Tyr
145 150 155 160

ttc cat gat ttc ttt aac tat gcc gga atc cat cgc agc gta atg ctc 528
Phe His Asp Phe Phe Asn Tyr Ala Gly Ile His Arg Ser Val Met Leu
165 170 175

tac acc acg ccg aac acc tgg gtg gac gat atc acc gtg gtg acg cat 576

```

Tyr	Thr	Thr	Pro	Asn	Thr	Trp	Val	Asp	Asp	Ile	Thr	Val	Val	Thr	His	
			180					185					190			
gtc	gcg	caa	gac	tgt	aac	cac	gcg	tct	gtt	gac	tgg	cag	gtg	gtg	gcc	624
Val	Ala	Gln	Asp	Cys	Asn	His	Ala	Ser	Val	Asp	Trp	Gln	Val	Val	Ala	
		195					200					205				
aat	ggc	gat	gtc	agc	gtt	gaa	ctg	cgt	gat	gcg	gat	caa	cag	gtg	gtt	672
Asn	Gly	Asp	Val	Ser	Val	Glu	Leu	Arg	Asp	Ala	Asp	Gln	Gln	Val	Val	
	210					215					220					
gca	act	gga	caa	ggc	act	agc	ggg	act	ttg	caa	gtg	gtg	aat	ccg	cac	720
Ala	Thr	Gly	Gln	Gly	Thr	Ser	Gly	Thr	Leu	Gln	Val	Val	Asn	Pro	His	
225					230					235					240	
ctc	tgg	caa	ccg	ggc	gaa	ggc	tat	ctc	tat	gaa	ctg	tgc	gtc	aca	gcc	768
Leu	Trp	Gln	Pro	Gly	Glu	Gly	Tyr	Leu	Tyr	Glu	Leu	Cys	Val	Thr	Ala	
				245					250					255		
aaa	agc	cag	aca	gag	tgt	gat	atc	tac	ccg	ctt	cgc	gtc	ggc	atc	cgg	816
Lys	Ser	Gln	Thr	Glu	Cys	Asp	Ile	Tyr	Pro	Leu	Arg	Val	Gly	Ile	Arg	
			260					265					270			
tca	gtg	gca	gtg	aag	ggc	gaa	cag	ttc	ctg	att	aac	cac	aaa	ccg	ttc	864
Ser	Val	Ala	Val	Lys	Gly	Glu	Gln	Phe	Leu	Ile	Asn	His	Lys	Pro	Phe	
		275					280					285				
tac	ttt	act	ggc	ttt	ggc	cgt	cat	gaa	gat	gcg	gac	ttg	cgt	ggc	aaa	912
Tyr	Phe	Thr	Gly	Phe	Gly	Arg	His	Glu	Asp	Ala	Asp	Leu	Arg	Gly	Lys	
	290					295					300					
gga	ttc	gat	aac	gtg	ctg	atg	gtg	cac	gac	cac	gca	tta	atg	gac	tgg	960
Gly	Phe	Asp	Asn	Val	Leu	Met	Val	His	Asp	His	Ala	Leu	Met	Asp	Trp	
305					310					315					320	
att	ggg	gcc	aac	tcc	tac	cgt	acc	tgc	cat	tac	cct	tac	gct	gaa	gag	1008
Ile	Gly	Ala	Asn	Ser	Tyr	Arg	Thr	Ser	His	Tyr	Pro	Tyr	Ala	Glu	Glu	
				325					330					335		
atg	ctc	gac	tgg	gca	gat	gaa	cat	ggc	atc	gtg	gtg	att	gat	gaa	act	1056
Met	Leu	Asp	Trp	Ala	Asp	Glu	His	Gly	Ile	Val	Val	Ile	Asp	Glu	Thr	
			340					345					350			
gct	gct	gtc	ggc	ttt	aac	ctc	tct	tta	ggc	att	ggc	ttc	gaa	gcg	ggc	1104
Ala	Ala	Val	Gly	Phe	Asn	Leu	Ser	Leu	Gly	Ile	Gly	Phe	Glu	Ala	Gly	
		355					360					365				
aac	aag	ccg	aaa	gaa	ctg	tac	agc	gaa	gag	gca	gtc	aac	ggg	gaa	act	1152
Asn	Lys	Pro	Lys	Glu	Leu	Tyr	Ser	Glu	Glu	Ala	Val	Asn	Gly	Glu	Thr	
	370					375					380					
cag	caa	gcg	cac	tta	cag	gcg	att	aaa	gag	ctg	ata	gcg	cgt	gac	aaa	1200
Gln	Gln	Ala	His	Leu	Gln	Ala	Ile	Lys	Glu	Leu	Ile	Ala	Arg	Asp	Lys	
385					390					395					400	
aac	cac	cca	agc	gtg	gtg	atg	tgg	agt	att	gcc	aac	gaa	ccg	gat	acc	1248
Asn	His	Pro	Ser	Val	Val	Met	Trp	Ser	Ile	Ala	Asn	Glu	Pro	Asp	Thr	
				405					410					415		
cgt	ccg	caa	ggc	gca	ccg	gaa	tat	ttc	gcg	cca	ctg	gcg	gaa	gca	acg	1296
Arg	Pro	Gln	Gly	Ala	Arg	Glu	Tyr	Phe	Ala	Pro	Leu	Ala	Glu	Ala	Thr	
			420					425					430			
cgt	aaa	ctc	gac	ccg	acg	cgt	ccg	atc	acc	tgc	gtc	aat	gta	atg	ttc	1344
Arg	Lys	Leu	Asp	Pro	Thr	Arg	Pro	Ile	Thr	Cys	Val	Asn	Val	Met	Phe	
		435					440					445				

tgc	gac	gct	cac	acc	gat	acc	atc	agc	gat	ctc	ttt	gat	gtg	ctg	tgc	1392
Cys	Asp	Ala	His	Thr	Asp	Thr	Ile	Ser	Asp	Leu	Phe	Asp	Val	Leu	Cys	
	450					455					460					
ctg	aac	cgt	tat	tac	gga	tgg	tat	gtc	caa	agc	ggc	gat	ttg	gaa	acg	1440
Leu	Asn	Arg	Tyr	Tyr	Gly	Trp	Tyr	Val	Gln	Ser	Gly	Asp	Leu	Glu	Thr	
465					470					475					480	
gca	gag	aag	gta	ctg	gaa	aaa	gaa	ctt	ctg	gcc	tgg	cag	gag	aaa	ctg	1488
Ala	Glu	Lys	Val	Leu	Glu	Lys	Glu	Leu	Leu	Ala	Trp	Gln	Glu	Lys	Leu	
				485					490					495		
cat	cag	ccg	att	atc	atc	acc	gaa	tac	ggc	gtg	gat	acg	tta	gcc	ggg	1536
His	Gln	Pro	Ile	Ile	Ile	Thr	Glu	Tyr	Gly	Val	Asp	Thr	Leu	Ala	Gly	
			500					505					510			
ctg	cac	tca	atg	tac	acc	gac	atg	tgg	agt	gaa	gag	tat	cag	tgt	gca	1584
Leu	His	Ser	Met	Tyr	Thr	Asp	Met	Trp	Ser	Glu	Glu	Tyr	Gln	Cys	Ala	
		515					520					525				
tgg	ctg	gat	atg	tat	cac	cgc	gtc	ttt	gat	cgc	gtc	agc	gcc	gtc	gtc	1632
Trp	Leu	Asp	Met	Tyr	His	Arg	Val	Phe	Asp	Arg	Val	Ser	Ala	Val	Val	
	530					535					540					
ggt	gaa	cag	gta	tgg	aat	ttc	gcc	gat	ttt	gcg	acc	tcg	caa	ggc	ata	1680
Gly	Glu	Gln	Val	Trp	Asn	Phe	Ala	Asp	Phe	Ala	Thr	Ser	Gln	Gly	Ile	
545					550					555					560	
ttg	cgc	gtt	ggc	ggt	aac	aag	aaa	ggg	atc	ttc	act	cgc	gac	cgc	aaa	1728
Leu	Arg	Val	Gly	Gly	Asn	Lys	Lys	Gly	Ile	Phe	Thr	Arg	Asp	Arg	Lys	
			565						570					575		
ccg	aag	tcg	gcg	gct	ttt	ctg	ctg	caa	aaa	cgc	tgg	act	ggc	atg	aac	1776
Pro	Lys	Ser	Ala	Ala	Phe	Leu	Leu	Gln	Lys	Arg	Trp	Thr	Gly	Met	Asn	
			580					585					590			
ttc	ggt	gaa	aaa	ccg	cag	cag	gga	ggc	aaa	caa	tga					1812
Phe	Gly	Glu	Lys	Pro	Gln	Gln	Gly	Gly	Lys	Gln	*					
		595					600									

<210> 106  
 <211> 603  
 <212> PRT  
 <213> Escherichia coli

<300>  
 <308> Genbank #S69414  
 <309> 1994-09-23

<400> 106  
 Met Leu Arg Pro Val Glu Thr Pro Thr Arg Glu Ile Lys Lys Leu Asp  
 1 5 10 15  
 Gly Leu Trp Ala Phe Ser Leu Asp Arg Glu Asn Cys Gly Ile Asp Gln  
 20 25 30  
 Arg Trp Trp Glu Ser Ala Leu Gln Glu Ser Arg Ala Ile Ala Val Pro  
 35 40 45  
 Gly Ser Phe Asn Asp Gln Phe Ala Asp Ala Asp Ile Arg Asn Tyr Ala  
 50 55 60  
 Gly Asn Val Trp Tyr Gln Arg Glu Val Phe Ile Pro Lys Gly Trp Ala  
 65 70 75 80  
 Gly Gln Arg Ile Val Leu Arg Phe Asp Ala Val Thr His Tyr Gly Lys  
 85 90 95  
 Val Trp Val Asn Asn Gln Glu Val Met Glu His Gln Gly Gly Tyr Thr  
 100 105 110  
 Pro Phe Glu Ala Asp Val Thr Pro Tyr Val Ile Ala Gly Lys Ser Val  
 115 120 125

Arg	Ile	Thr	Val	Cys	Val	Asn	Asn	Glu	Leu	Asn	Trp	Gln	Thr	Ile	Pro
	130					135					140				
Pro	Gly	Met	Val	Ile	Thr	Asp	Glu	Asn	Gly	Lys	Lys	Lys	Gln	Ser	Tyr
145					150					155					160
Phe	His	Asp	Phe	Phe	Asn	Tyr	Ala	Gly	Ile	His	Arg	Ser	Val	Met	Leu
			165					170						175	
Tyr	Thr	Thr	Pro	Asn	Thr	Trp	Val	Asp	Ile	Thr	Val	Val	Thr	His	
			180					185				190			
Val	Ala	Gln	Asp	Cys	Asn	His	Ala	Ser	Val	Asp	Trp	Gln	Val	Val	Ala
		195					200					205			
Asn	Gly	Asp	Val	Ser	Val	Glu	Leu	Arg	Asp	Ala	Asp	Gln	Gln	Val	Val
	210					215				220					
Ala	Thr	Gly	Gln	Gly	Thr	Ser	Gly	Thr	Leu	Gln	Val	Val	Asn	Pro	His
225					230					235					240
Leu	Trp	Gln	Pro	Gly	Glu	Gly	Tyr	Leu	Tyr	Glu	Leu	Cys	Val	Thr	Ala
				245				250						255	
Lys	Ser	Gln	Thr	Glu	Cys	Asp	Ile	Tyr	Pro	Leu	Arg	Val	Gly	Ile	Arg
			260					265					270		
Ser	Val	Ala	Val	Lys	Gly	Glu	Gln	Phe	Leu	Ile	Asn	His	Lys	Pro	Phe
		275					280					285			
Tyr	Phe	Thr	Gly	Phe	Gly	Arg	His	Glu	Asp	Ala	Asp	Leu	Arg	Gly	Lys
	290					295				300					
Gly	Phe	Asp	Asn	Val	Leu	Met	Val	His	Asp	His	Ala	Leu	Met	Asp	Trp
305					310					315					320
Ile	Gly	Ala	Asn	Ser	Tyr	Arg	Thr	Ser	His	Tyr	Pro	Tyr	Ala	Glu	Glu
				325					330					335	
Met	Leu	Asp	Trp	Ala	Asp	Glu	His	Gly	Ile	Val	Val	Ile	Asp	Glu	Thr
			340					345					350		
Ala	Ala	Val	Gly	Phe	Asn	Leu	Ser	Leu	Gly	Ile	Gly	Phe	Glu	Ala	Gly
		355					360					365			
Asn	Lys	Pro	Lys	Glu	Leu	Tyr	Ser	Glu	Glu	Ala	Val	Asn	Gly	Glu	Thr
	370					375					380				
Gln	Gln	Ala	His	Leu	Gln	Ala	Ile	Lys	Glu	Leu	Ile	Ala	Arg	Asp	Lys
385					390					395					400
Asn	His	Pro	Ser	Val	Val	Met	Trp	Ser	Ile	Ala	Asn	Glu	Pro	Asp	Thr
				405					410					415	
Arg	Pro	Gln	Gly	Ala	Arg	Glu	Tyr	Phe	Ala	Pro	Leu	Ala	Glu	Ala	Thr
			420					425					430		
Arg	Lys	Leu	Asp	Pro	Thr	Arg	Pro	Ile	Thr	Cys	Val	Asn	Val	Met	Phe
		435					440					445			
Cys	Asp	Ala	His	Thr	Asp	Thr	Ile	Ser	Asp	Leu	Phe	Asp	Val	Leu	Cys
	450					455					460				
Leu	Asn	Arg	Tyr	Tyr	Gly	Trp	Tyr	Val	Gln	Ser	Gly	Asp	Leu	Glu	Thr
465					470					475					480
Ala	Glu	Lys	Val	Leu	Glu	Lys	Glu	Leu	Leu	Ala	Trp	Gln	Glu	Lys	Leu
				485					490					495	
His	Gln	Pro	Ile	Ile	Ile	Thr	Glu	Tyr	Gly	Val	Asp	Thr	Leu	Ala	Gly
			500					505					510		
Leu	His	Ser	Met	Tyr	Thr	Asp	Met	Trp	Ser	Glu	Glu	Tyr	Gln	Cys	Ala
		515					520					525			
Trp	Leu	Asp	Met	Tyr	His	Arg	Val	Phe	Asp	Arg	Val	Ser	Ala	Val	Val
	530					535					540				
Gly	Glu	Gln	Val	Trp	Asn	Phe	Ala	Asp	Phe	Ala	Thr	Ser	Gln	Gly	Ile
545					550					555					560
Leu	Arg	Val	Gly	Gly	Asn	Lys	Lys	Gly	Ile	Phe	Thr	Arg	Asp	Arg	Lys
				565					570					575	
Pro	Lys	Ser	Ala	Ala	Phe	Leu	Leu	Gln	Lys	Arg	Trp	Thr	Gly	Met	Asn
			580					585					590		
Phe	Gly	Glu	Lys	Pro	Gln	Gln	Gly	Gly	Lys	Gln					
		595					600								

<210> 107

<211> 277

<212> DNA

<213> Artificial Sequence



<220>  
<223> Nopaline Synthase Terminator Sequence

<300>  
<308> U09365  
<309> 1995-10-17

```
<400> 107
gagctcgaat ttccccgac gttcaaacat ttggcaataa agtttcttaa gattgaatcc 60
tggtgcccgt cttgcgatga ttatcatata atttctgttg aattacgtta agcatgtaat 120
aattaacatg taatgcatga cgttatttat gagatgggtt tttatgatta gagtcccgca 180
attatacatt taatagcgga tagaaaacaa aatatagcgc gcaaactagg ataaattatc 240
gcgcgcggtg tcactctatg tactagatcg ggaattc 277
```

<210> 108  
<211> 3451  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> HindIII Fragment containing the beta-glucuronidase coding sequence, the rDNA intergenic spacer, and the Mast1 sequence

```
<400> 108
aagcttgacc tggaatatcg cgagtaaact gaaaatcacg gaaaatgaga aatacacact 60
ttaggacgtg aaatatggcg aggaaaactg aaaaagggtg aaaatttaga aatgtccact 120
gtaggacgtg gaatatggca agaaaactga aaatcatgga aaatgagaaa catccacttg 180
acgacttgaa aaatgacgaa atcactaaaa aacgtgaaaa atgagaaatg cactactgaag 240
gactccgctg gaattcgatt gtgctagcca atgtttaaca agatgtcaag cacaatgaat 300
gttggtggtt ggtggtcgtg gctggcggtg gtggaaaatt gcggtggttc gagcggtagt 360
gatcgcgcat ggttggtggt tgcagcggtg tttgatatcg gaatcactta tgggtggtgt 420
cacaatggag gtgcgtcatg gttattgggt gttggtcatc tatatatatt tatataata 480
ttaagtattt tacctatttt ttacatattt tttttaaatt ttatgcattg tttgtatttt 540
taaatagttt ttatcgtact tgttttataa aatattttat tattttatgt gttattattt 600
tacttgatgt attggaaatt ttctccattg ttttttctat atttataata attttcttat 660
ttttttttgt tttattatgt attttttctg tttataataa atattttatta aaaaaaatat 720
tattttttga aaatatatca tttacaatgt ttaaaagtca tttgtgaata tattagctaa 780
gttgtaactt tttgtgtgca tttgtgtgtg tacatgtcta ttatgattct ctggccaaaa 840
catgtctact cctgtcactt ggggtttttt ttttaagaca taatcactag tgatttatatc 900
tagactgaag gcgggaaacg acaatctgat catgagcggg gaattaaggg agtcacgtta 960
tgaccccgcg cgatgacgcg ggacaagcgg ggaactgaca gaaccgcaac 1020
gttgaaggag ccactcagcc gcgggtttct ggagtttaat gagctaagca catacgtcag 1080
aaaccattat tgcgcgttca aaagtcgctt aaggtcacta tcagctagca aatatttctt 1140
gtcaaaaaat gttccataaa ttcccctcgg ttttacgttt tatccaatta gactctcata 1200
ttcactctca atccaaataa tctgcaccgg atctcgagat cgaattcccc cggccgcgaa 1260
ttcactagtg gatccccggg tacggtcagt cccttatggt acgtcctgta gaaaccccaa 1320
cccgtgaaat caaaaaactc gacggcctgt gggcattcag tctggatcgc gaaaactgtg 1380
gaattgagca gcgttggtgg gaaagcgcgt tacaagaaag ccgggcaatt gctgtgccag 1440
gcagttttaa cgatcagttc gccgatgcag atattcgtaa ttatgtgggc aacgtctggt 1500
atcagcgcca agtctttata ccgaaagggt gggcaggcca cgtatcgtg ctgcgtttcg 1560
atgcggtcac tcattacggc aaagtgtggg tcaataatca ggaagtgatg gagcatcagg 1620
gcggtatcac gccatttgaa gccgatgtca cgccgatgtg tattgcccgg aaaagtgtac 1680
gtatcacagt ttgtgtgaac aacgaactga actggcagac tatcccgcgg ggaatggtga 1740
ttaccgacga aaacggcaag aaaaagcagt cttactcca tgatttcttt aactacgccg 1800
ggatccatcg cagcgtaatg ctctacacca cgccgaacac ctgggtggac gatatacccg 1860
tggtgacgca tgtcgcgcaa gactgttaacc acgcgtctgt tgactggcag gtgtgggcca 1920
atggtgatgt cagcgttgaa ctgctgatg cggatcaaca ggtggttgca actggacaag 1980
gcaccagcgg gactttgcaa gtggtgaatc cgcacctctg gcaaccgggt gaaggttatc 2040
tctatgaaat gtacgtcaca gctaaaagcc agacagagtg tgatatctac ccgctgcgag 2100
tcggcatccg gtcagtggca gtgaagggcg aacagttcct gatcaaccac aaaccgttct 2160
actttactgg cttttggcgt catgaagatg cggatttgcg cggcaaagga ttcgataacc 2220
tgctgatggt gcacgatcac gcattaatgg actggattgg ggccaactcc taccgtacct 2280
cgcattaccc ttacgctgaa gagatgctcg actgggcaga tgaacatggc atcgtggtga 2340
ttgatgaaac tgcagctgtc ggctttaacc tctctttagg catttggttc gaagcgggca 2400
acaagccgaa agaactgtac agcgaagagg ggaactcag caggcgact 2460
tacaggcgat taaagagctg atagcgcgtg acaaaaacca cccaagcgtg gtgatgtgga 2520
```

gtattgccaa	cgaaccggat	accggtccgc	aagggtgcacg	ggaatatttc	gcgccactgg	2580
cggaagcaac	gcgtaaactc	gatccgacgc	gtccgatcac	ctgcgtcaat	gtaatgttct	2640
gcgacgctca	caccgatacc	atcagcgatc	tctttgatgt	gctgtgcctg	aaccgttatt	2700
acggttggtg	tgcccaaagc	ggcgatttgg	aaacggcaga	gaaggtagctg	gaaaaagaac	2760
ttctggcctg	gcaggagaaa	ctgcatcagc	cgattatcat	caccgaatac	ggcgtggata	2820
cgtagccggg	gctgcactca	atgtacaccg	acatgtggag	tgaagagtat	cagtgtgcat	2880
ggctggatat	gtatcaccgc	gtctttgatc	gcgtcagcgc	cgtegtcggt	gaacagggtat	2940
ggaatttcgc	cgattttgcg	acctcgcaag	gcatattgcg	cgttggcggt	aacaagaagg	3000
ggatcttcac	ccgcgaccgc	aaaccgaagt	cggcggtctt	tctgctgcaa	aaacgctgga	3060
ctggcatgaa	cttcgggtgaa	aaaccgcagc	agggaggcaa	acaatgaatc	aacaactctc	3120
ctggcgcaacc	atcgctgggt	acagcctcgg	gaattgcgta	ccgagctcga	atttccccga	3180
tcgttcaaac	atattggcaat	aaagtcttct	aagattgaat	cctgttgccg	gtcttgcgat	3240
gattatcata	taattttctgt	tgaattacgt	taagcatgta	ataattaaca	tgtaatgcac	3300
gacgttatatt	atgagatggg	tttttatgat	tagagtcccg	caattataca	tttaatacgc	3360
gatagaaaac	aaaatatagc	gcgcaacta	ggataaatta	tcgcgcgcgg	tgatcatctat	3420
gttactagat	cgggaattcg	atatcaagct	t			3451

<210> 109  
 <211> 14627  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pAg11a Plasmid

<400> 109						
catgccaaac	acagggttcc	cctcgggatc	aaagtacttt	gateccaacc	ctccgctgct	60
atagtgcagt	cggttcttga	cgttcagtg	agccgtcttc	tgaaaacgac	atgtcgcaca	120
agtcctaagt	tacgcgacag	gctgcccgcc	tgcccttttc	ctggcggttt	cttgtcgctg	180
gttttagtgc	cataaagtag	aatacttgcg	actagaaccg	gagacattac	gccatgaaca	240
agagcgccgc	cgctggcctg	ctgggctatg	cccgcgtcag	caccgacgac	caggacttga	300
ccaaccaacg	ggccgaactg	cacgcggcgc	gctgcaccaa	gctgttttcc	gagaagatca	360
ccggcaccag	gcgcgaccgc	ccggagctgg	ccaggatgct	tgaccaccta	cgccctggcg	420
acgtttgtgac	agtgaccagg	ctagaccgcc	tgcccgcgag	caccgcgcgac	ctactggaca	480
ttgccgagcg	catccaggag	gccggcgccg	gcctgcgtag	cctggcagag	ccgtgggccc	540
acaccaccac	gccggccggc	cgcatggtgt	tgaccgtggt	cgccggcatt	gccgagttcg	600
agcgttccct	aatcatcgac	cgaccccgga	gcgggcgcga	ggccgcgaag	gcccaggagg	660
tgaagtttgg	ccccgccttc	accctcaccc	cggcacagat	cgcgcacgcc	cgcgagctga	720
tcgaccagga	aggccgcacc	gtgaaagagg	cggtgcact	gcttggcggt	catcgctcga	780
ccctgtaccg	cgcacttgag	cgcagcgagg	aagtgcagcc	caccgaggcc	aggcgccgcg	840
gtgccttccg	tgaggacgca	ttgaccgagg	ccgacgccct	ggcgggccgc	gagaatgaac	900
gccaagagga	acaagcatga	aaccgcacca	ggacggccag	gacgaaccgt	ttttcattac	960
cgaagagatc	gaggcggaga	tgatcgcggc	cgggtacgtg	ttcgagccgc	ccgcgcacgt	1020
ctcaaccgtg	cggttgcgat	aaatcctggc	cggtttgtct	gatgccaaag	tgggggccctg	1080
gcccggccag	ttggccgctg	aagaaaccga	gcgcgcgcgt	ctaaaaagggt	gatgtgtatt	1140
tgagtaaaac	agcttgcgct	atgcggctcg	tgcgatatatg	atgcgatgag	taataataca	1200
aatacgcaag	gggaacgcat	gaagggtatc	gctgtactta	accagaaagg	cgggtcaggc	1260
aagacgacca	tcgcaaccca	tctagcccg	gccctgcaac	tcgcccgggg	cgatgttctg	1320
ttagtgcatt	cogatcccca	gggcagtgcc	cgcgattggg	cgcccttacga	ggaagatcaa	1380
ccgctaaccg	ttgtcggcac	cgaccgcccg	acgattgacc	gcgacgtgaa	ggccatcggc	1440
cggcgcgact	tcgtagtgat	cgacggagcg	ccccaggcgg	cggaacttgg	tggttccgcg	1500
atcaaggcag	ccgacttcgt	gctgattccg	gtgcagccaa	gcccttacga	catatggggc	1560
accgcccagc	tggtggagct	ggttaagcag	cgcattgagg	tcacggatgg	aaggctacaa	1620
gcggcctttg	tcgtgtcgcg	ggcgatcaaa	ggcacgcgca	tcggcggtga	ggttgcggag	1680
gcgctggccg	ggtacgagct	gcccattctt	gagtcgccga	tcacgcagcg	cgtgagctac	1740
ccaggcactg	ccgcgcgcgg	cacaaccggt	cttgaatcag	aaccgagggg	cgacgctgcc	1800
cgcgaggtcc	aggcgctggc	cgctgaaatt	aaatcaaaac	tcatttgagt	taatgaggtg	1860
aagagaaaaa	gagcaaaaag	acaaacacgc	taagtgcggg	ccgtccgagc	gcacgcagca	1920
gcaaggctgc	aacgttggcc	agcctggcag	acacgcagcg	catgaagcgg	gtcaactttc	1980
agttgcggcg	ggaggatcac	accaagctga	agatgtacgc	ggtacgcgca	ggcaagacca	2040
ttaccgagct	gctatctgaa	tacatcgccg	agctaccaga	gtaaatgagc	aatgaataaa	2100
atgagtagat	gaatttttagc	ggctaaagga	ggcggtcatgg	aaaatcaaga	acaaccaggc	2160
accgaagccg	tggaatgcc	catgtgtgga	ggaacggggc	ggtggccagg	cgtaagcgcc	2220
tggttgtct	gccggccctg	caatggcact	ggaaccccca	agcccaggga	atcggtgga	2280
cggtcgcaaa	ccatccggcc	cggtacaaat	cgccgcggcg	ctgggtgatg	acctggtgga	2340
gaagttgaag	gccgcgcagg	ccgccagcg	gcaacgcac	gaggcagaag	cacgcccggg	2400
tgaatcgctg	caagcggccg	ctgatcgaat	ccgcaagaa	tcccggcaac	cgccggcagg	2460

cggtgcgccg	tgcattagga	agccgcccga	gggagcagag	caaccagatt	ttttcgttcc	2520
gatgctctat	gacgtgggca	cccgcgatag	tgcagcctc	atggacgtgg	ccgttttccg	2580
tctgtcgaag	cgtgaccgac	gagctggcga	ggtgatccgc	tacgagcttc	cagacgggca	2640
cgtagagggt	tccgcagggc	cgcccgccat	ggccagtggt	tgggattacg	acctgggtact	2700
gatggcggtt	tcccatctaa	ccgaatccat	gaaccgatac	cggaaggga	agggagacaa	2760
gcccggccgc	gtgttccgtc	cacacgttgc	ggacgtactc	aagttctgcc	ggcgagccga	2820
tggcggaag	cagaaagacg	acctggtaga	aacctgcatt	cggttaaaca	ccacgcacgt	2880
tgccatgcag	cgtacgaaga	aggccaagaa	cgccgcctcg	gtgacgggtat	ccgaggggtga	2940
agccttgatt	agccgctaca	agatcgtaaa	gagcgaaacc	gggcggcccg	agtacatcga	3000
gatcgagcta	gctgatttga	tgtaccgcga	gatcacagaa	ggcaagaacc	cggacgtgct	3060
gacggttcac	cccgattact	ttttgatcga	tcccgccatc	ggccggtttc	tctaccgctt	3120
ggcacgccgc	gcccagggca	aggcagaagc	cagatgggtg	ttcaagacga	tctacgaacg	3180
cagtgccagc	gcccggagagt	tcaagaagtt	ctgtttcacc	gtgcgcaagc	tgatcgggtc	3240
aaatgacctg	ccggagtacg	atttgaagga	ggaggcgggg	caggctggcc	cgatcctagt	3300
catgcgctac	cgcaacctga	tgcagggcga	agcatccgcc	ggttcctaata	gtacggagca	3360
gatgctaggg	caaattggcc	tagcagggga	aaaagggtcga	aaagggtctct	ttcctgtgga	3420
tagcacgtac	attgggaacc	caaagccgta	cattgggaac	cggaaccctg	acattgggaa	3480
cccaaagccg	tacattggga	accggtcaca	catgtaagtg	actgatataa	aagagaaaaa	3540
aggcgatttt	tccgcctaaa	actctttaaa	acttattaaa	actcttaaaa	cccgcctggc	3600
ctgtgcataa	ctgtctggcc	agcgacacgc	cgaagagctg	caaaaagcgc	ctaccgctcg	3660
gtcgtgcgc	tccctacgcc	ccgcgccttc	gcgctggcct	atcgcgcccg	ctggccgctc	3720
aaaaatggct	ggcctacggc	caggcaatct	accagggcgc	ggacaagccg	cgccgtcgcc	3780
actcgaccgc	cggcgccac	atcaaggcac	cctgcctcgc	gcggttcggg	gatgacgggtg	3840
aaaacctctg	acacatgcag	ctcccggaga	cggctcacgc	ttgtctgtaa	gcggtatgccg	3900
ggagcagaca	agcccgctcag	ggcgcgctcag	cggttgctgg	cggttgctgg	ggcgagcca	3960
tgaccagctc	acgtagcgat	agcggagtgt	atactggctt	aactatgcgg	catcagagca	4020
gattgtactg	agagtgcacc	atatgcgggtg	tgaataaccg	cacagatgcg	taaggagaaa	4080
ataccgcctc	aggcgctctt	ccgcttccctc	gctcactgac	tcgctgcgct	cggtcggttcg	4140
gctgcgccga	gcggtatcag	ctcactcaaa	ggcggttaata	cggttatcca	cagaatcagg	4200
ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	4260
ggccgcggtg	ctggcggtttt	tccataggct	ccgccccctt	gacgagcatc	acaaaaatcg	4320
acgctcaagt	cagagggtggc	gaaaccgcac	aggactataa	agataccagg	cggttcccc	4380
tggaaagctc	ctcgtgcgct	ctcctgttcc	gaccctgccc	cttaccggat	acctgtccgc	4440
ctttctccct	tccgggaagcg	tggcgctttc	tcatagctca	cgctgtagggt	atctcagttc	4500
gggtgtaggtc	ggttcgctcca	agctgggctg	tgtgcacgaa	ccccccgctt	agcccgagcc	4560
ctgcgcctta	tccggtaact	atcgctttga	gtccaaaccg	gtaagacacg	acttatcgcc	4620
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	4680
gttcttgaag	tggtggccta	actacggcta	cactgaagg	acagtatttg	gtatctgcgc	4740
tctgctgaag	ccagttacct	tccgaaaaag	agttggtagc	tcttgatccg	gcaaaacaa	4800
caccgctgggt	agcgggtgggt	tttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	4860
atctcaagaa	gatcccttga	tcttttctac	ggggtctgac	gctcagtgga	acgaaaactc	4920
acgttaagggt	atttttggtca	tgcattctag	gtactataaa	aattcatcca	gtaaaaatata	4980
atattttatt	ttctcccaat	caggcttgat	ccccagtaag	tcaaaaaata	gctcgacata	5040
ctgtttcttc	ccgatatact	ccctgatcga	ccggacgcag	aaggcaatgt	cataccactt	5100
gtccgcctcg	cgcttctcc	caagatcaat	aaagccactt	actttgcca	ctttcacaaa	5160
gatgttgctg	tctcccagggt	cgccgtggga	aaagacaagt	tccctcttcg	gcttttccgt	5220
ctttaaaaaa	tcatacagct	cgccgggatc	tttaaatgga	gtgtcttctt	cccagttttc	5280
gcaatccaca	tccggccagat	cggtattcag	taagtaatcc	aattcgggta	agcggctgtc	5340
taagctattc	gtataggggac	aatccgatat	gtcgatggag	tgaaagagcc	tgatgcactc	5400
cgcatacagc	tcgataatct	tttcagggtc	ttgttcatct	tcatactctt	ccgagcaaa	5460
gacgccatcg	gcctcactca	tgagcagatt	gctccagcca	tcatgccgtt	caaagtgcag	5520
gacctttgga	acaggcagct	ttccttccag	ccatagcatc	atgtcctttt	cccgttccac	5580
atcatagggtg	gtccctttat	accggctgtc	cgctattttt	aaatatagggt	tttcattttc	5640
tcccaccagc	ttatatacct	tagcaggaga	cattccttcc	gtatctttta	cgagcggtta	5700
tttttcgatac	agttttttca	attccgggtga	tattctcatt	ttagccattt	attatttccct	5760
tccctttttc	tacagttatt	aaagataccc	caagaagcta	attataacaa	gacgaactcc	5820
aattcactgt	tccctgcat	ctaaaacctt	aaataccaga	aaacagcttt	ttcaaagtgt	5880
ttttcaaagt	tggcggtataa	catagtatcg	acggagccga	ttttgaaacc	gcggtgatca	5940
caggcagcaa	cgctctgtca	tcgttacaat	caacatgcta	ccctccgcca	gatcatccgt	6000
gtttcaaaac	cggcagctta	gltgcccgtt	ttccgaatag	catcggtaac	atgagcaaa	6060
tctgcccgcct	tacaacggct	ctcccgcgtga	cgccgtcccg	gactgatggg	ctgcctgtat	6120
cgaattgggtga	ttttgtgccc	agctgcccgt	cggggagctg	ttggctgggt	gggtggcagga	6180
tatatgtgtg	tgtaaacaaa	ttgacgctta	gacaacttaa	taacacattg	cggagctttt	6240
taatgtactg	aattaaacgcc	gaattaattc	gggggatctg	gatttttagta	ctggattttg	6300
gtttttaggaa	ttagaaattt	tattgataga	agttattttac	aaatacaaat	acataactag	6360
ggtttcttat	atgctcaaca	catgagcgaa	accttatagg	aaccttaatt	cccttatctg	6420
ggaactactc	acacattatt	atggagaaa	tcgagtcaaa	tctcggtgac	gggcaggacc	6480

ggacggggcg	gtaccggcag	gctgaagtcc	agctgccaga	aaccacgctc	atgccagttc	6540
ccgtgcttga	agccggccgc	ccgcagcatg	ccgcgggggg	catatccgag	cgccctcgtc	6600
atgcgacgcg	tcgggtcggt	gggcagcccg	atgacagcga	ccacgctctt	gaagccctgt	6660
gcctccaggg	acttcagcag	gtgggtgtag	agcgtggagc	ccagtcocgt	ccgctgggtg	6720
cggggggaga	cgtaacgggt	cgactcggcc	gtccagtcgt	aggcgttgcg	tgccctccag	6780
gggcccgcgt	aggcgatgcc	ggcgacctcg	ccgtccacct	cggcgacgag	ccagggatag	6840
cgctcccga	gacggacgag	gtcgtccgtc	cactcctgcg	gttcctgcg	ctcggtagcg	6900
aagttgaccg	tgcttgtctc	gatgtagtgg	ttgacgatgg	tgacagccgc	cggcaggtcc	6960
gcctcggtgg	cacggcggat	gtcggccggg	cgctcgttct	ggctcatggg	agactcgaga	7020
gagatagatt	tgtagagaga	gactgggtgat	ttcagcgtgt	cctctccaaa	tgaaatgaac	7080
ttccttatat	agagggaagg	cttgcggaagg	atagtgggat	tgtgcgtcat	cccttacgtc	7140
agtggagata	tcacatcaat	ccacttgctt	tgaagacgtg	gttggaaacgt	cttctttttc	7200
cacgatgctc	ctcgtgggtg	gggggtccatc	tttgggacca	ctgtcggcag	aggcatcttg	7260
aacgatagcc	tttcctttat	cgcaatgatg	cgatttggtag	gtgccacctt	ccttttctac	7320
tgctcttttg	atgaagtgc	agatagctgg	gcaatggaat	ccgaggaggt	ttcccgatat	7380
taccctttgt	tgaaaagtct	caatagccct	ttggctttct	gagactgtat	ctttgatatt	7440
cttggagtag	acgagagtgt	cggtctccac	catgttatca	catcaatcca	cttgctttga	7500
agacgtgggt	ggaaagctct	ctttttccac	gatgctcctc	gtgggtgggg	gtccatcttt	7560
gggaccactg	tcggcagagg	catcttgaa	catagccctt	ccttttatcgc	aatgatggca	7620
tttgtagggt	ccacctctct	ttctacttgt	ccttttgatg	aagtgcagca	tagctgggca	7680
atggaatccg	aggaggtttc	ccgatattac	cctttgttga	aaagtctcaa	tagccctttg	7740
gtcttctgag	actgtatctt	tgatattctt	ggagtagacg	agagtgtcgt	gctccaccat	7800
ggttgcaagc	tgctctagcc	aatacgcaaa	ccgctctctc	ccgcgcgttg	gccgattcat	7860
taatgcagct	ggcacgcacg	gtttcccgac	tggaagcgg	gcagtgcg	caacgcaatt	7920
aatgtgagtt	agctcactca	ttaggcaccc	caggctttac	actttatgct	tccggctcgt	7980
atggtgtgtg	gaattgtgag	cggaatacaa	cggtatacag	gaaacagcta	tgaccatgat	8040
tacgaattcg	agccttgact	agagggtcga	tgaaaaaaat	acatgataag	atacattgat	8100
gagtttgagc	aaaccacaac	tagaatgcag	agctgcaata	gctttatttg	tgaaatttgt	8160
gatgctattg	cttttatttg	aaccattata	gatcatccag	aacaagtgg	ggtgggcgaa	8220
gaactccagc	atgagatccc	cgcgctggag	cggtatccag	ccggcgctcc	ggaaaacgat	8280
tccgaagccc	aacctttcat	agaaggcggc	ggtggaatcg	aaatctcgta	gcacgtgtca	8340
gtcctgctcc	tcggccacga	agtgcacgca	gttgccggcc	gggtcgcgca	gggcgaactc	8400
ccgccccac	ggctgctcgc	cgatctcggt	cagctcgtcc	ccggaggcgt	cccgaagt	8460
cgtggacacg	acctccgacc	actcggcgta	ctggaccg	aggccgcgca	cccacacca	8520
ggccagggtg	ttgtccggca	ccacctggtc	ctggagcgc	ctgatgaaca	gggtcacgtc	8580
gtcccggacc	acaccggcga	agtcgtcctc	cacgaagtcc	cgggagaacc	cgagccggtc	8640
gggtccagaac	tcgaccgctc	cggcgacgtc	gcgcgcggtg	agcaccggaa	cggcactggt	8700
caacttggcc	atggatccag	atttcgctca	agttagtata	aaaaagcagg	cttcaatcct	8760
cgaggaattc	gatcgacact	ctcgtctact	ccaagaatat	caaaagataca	gtctcagaag	8820
accaaagggc	tattgagact	tttcaacaaa	gggtaatatc	gggaaacctc	ctcggattcc	8880
atggccagc	tatctgtcac	ttcatcaaaa	ggacagtaga	aaaggaaggt	ggcacctaca	8940
aatgccatca	ttgcgataaa	ggaaaaggcta	tcgttcaaga	tgccctctg	gacagtgggtc	9000
ccaaagatgg	acccccaccc	acgaggagca	tcgtggaaaa	agaagacgtt	ccaaccacgt	9060
cttcaaagca	agtggattga	tgtgataaca	ttggtggagca	cgacactctc	gtctactcca	9120
agaatatcaa	agatacagtc	tcagaagacc	aaagggctat	tgagactttt	caacaaaagg	9180
taatatcggt	aaacctcctc	ggattccatt	gcccagctat	ctgtcacttc	atcaaaaagga	9240
cagtagaaaa	ggaaggtggc	acctacaaat	gccatcattg	cgataaagga	aaggctatcg	9300
ttcaagatgc	ctctgccgac	agtgggtcca	aagatggacc	cccacccacg	aggagcatcg	9360
tggaaaaaaga	agacgttcca	accacgtcct	caaagcaagt	ggattgatgt	gatattctcca	9420
ctgacgtaag	ggatgacgca	caatcccact	atccttcgca	agaccttctc	ctatataagg	9480
aagttcattt	catttggaga	ggacacgctg	aaatcaccag	tctctctcta	caaactctatc	9540
tctctcgagc	tttcgcagat	ccgggggggg	aatgagatat	gaaaaagcct	gaactcaccg	9600
cgacgtctgt	cgagaagttt	ctgatcgaaa	agttcgacag	cgtctccgac	ctgatgcagc	9660
tctcgagggg	cgaagaatct	cgtgctttca	gcttcgatgt	aggagggcgt	ggatatgtcc	9720
tgccgggtaaa	tagctgcgcc	gatgggtttc	acaaagatcg	ttatgtttat	cggcactttg	9780
catcgcccg	gctcccgaat	ccggaaagtgc	ttgacattgg	ggagtttagc	gagagcctga	9840
cctattgcat	ctcccgcgt	gcacagggtg	tcacgttgca	agacctgcct	gaaaccgaac	9900
tgcccgctgt	tctacaaccg	gtcgcggagg	ctatggatgc	gatcgctg	gccgatctta	9960
gccagacgag	cgggttcggc	ccattcggag	cgcaaggaaat	cggtcaatac	actacatggc	10020
gtgatttcat	atgcgcgatt	gctgatcccc	atgtgtatca	ctggcaaaact	gtgatggacg	10080
acaccgtcag	tgcttcgctc	gcgcaggttc	tcgatgagct	gatgctttgg	gccgaggact	10140
gccccgaagt	ccggcacctc	gtgcacgcgg	atcttcggctc	caacaatgtc	ctgacggaca	10200
atggccgcat	aacagcggtc	attgactgga	cgaggcgcat	gttcggggat	tcccaatacg	10260
aggtcgccaa	catcttcttc	tgaggggcgt	ggttggcttg	tatggagcag	cagacgcgct	10320
acttcgagcg	gaggcatccg	gagcttgcag	gatcgccacg	actccggg	tatatgctcc	10380
gcatttggct	tgaccaactc	tatcagagct	tgttgacgg	caatttctgat	gatgcagctt	10440
gggcgcaggg	tcgatgcgac	gcaatcgtcc	gatccggagc	cgggactgtc	gggcgtacac	10500

aaatcgcccc	cagaagcgcg	gccgtctgga	ccgatggctg	tgtagaagta	ctcgccgata	10560
gtggaaacccg	acgccccagc	actcgctccga	gggcaaagaa	atagagtaga	tgccgaccgg	10620
atctgtcgat	cgacaagctc	gagtttctcc	ataataatgt	gtgagtagtt	cccagataag	10680
ggaattaggg	ttcctatagg	gtttcgctca	tgtgttgagc	atataagaaa	cccttagtat	10740
gtatttgat	ttgtaaaata	cttctatcaa	taaaatttct	aattcctaaa	acaaaaatcc	10800
agtactaaaa	tccagatccc	ccgaattaat	tccgctgtaa	ttcagatcaa	gcttgacctg	10860
gaatatcgcg	agtaaactga	aaatcacgga	aaatgagaaa	tacacacttt	aggacgtgaa	10920
atatggcgag	gaaaaactgaa	aaagggtggaa	aatttagaaa	tgtccactgt	aggacgtgga	10980
atatggcaag	aaaactgaaa	atcatggaaa	atgagaaaca	tccacttgac	gacttgaaaa	11040
atgacgaaat	cactaaaaaa	cgtgaaaaat	gagaaatgca	cactgaagga	ctccgcgagg	11100
attcgattgt	gctagccaat	gtttaaccaag	atgtcaagca	caatgaatgt	tggtggttgg	11160
tggtcggtgg	tggtcggtgg	ggaaaaattgc	gggtggttcca	gcggtagtga	tcggcgatgg	11220
ttggtgtttg	cagcggtgtt	tgatatcgga	atcacattatg	gtggttgtca	caatggagg	11280
gcgtcatgtt	tattggtggt	tggtcatcta	tatatatttta	taataatatt	aagtatttta	11340
cctatttttt	acataattttt	tattaaattt	atgcattgtt	tgtattttta	aatagttttt	11400
atcgactttg	ttttataaaa	tattttatta	ttttatgtgt	tatattatta	cttgatgtat	11460
tggaattttt	ctccattgtt	ttttctatat	ttataataat	ttttctattt	tttttgtttt	11520
tattatgtat	tttttcgttt	tataataaat	atattattaaa	aaaaatatta	tttttgtaaa	11580
atatatcatt	tacaattgtt	aaaagtcatt	tgtgaatata	ttagctaagt	tgtacttctt	11640
tttgtgcatt	tggtgttgta	catgtcttatt	atgattctct	ggccaaaaca	tgctactctc	11700
tgtcacttgg	gttttttttt	tttaagacata	atcactagt	attatatcta	gactgaaggc	11760
gggaaacgac	aatctgatca	tgagcggaga	attaagggag	tcacgttatg	acccccgccc	11820
atgacgcggg	acaagccgtt	ttacgttttg	aactgcagaca	accgcaacgt	tgaaggagcc	11880
actcagccgc	gggtttctgg	agtttaatga	gctaagcaca	tacgtcagaa	accattattg	11940
cgcggttcaaa	agtcgcctaa	ggtcactatc	agctagcaaa	tatttcttgt	caaaaatgct	12000
ccactgacgt	tccataaatt	cccctcggt	tccaattaga	gtctcatatt	cactctcaat	12060
ccaaataatc	tgacccggat	ctcgagatcg	aattcccgcg	gccgcgaatt	cactagtggg	12120
tccccgggta	cggtcagtc	cttatgtttac	gtcctgtaga	aaccccaacc	cgtgaaatca	12180
aaaaactcga	cggcctgtgg	gcattcagtc	tgcctgcgca	aaactgtgga	attgagcagc	12240
gttggtggga	aagcgcgtta	caagaaagcc	gggcaattgc	tgtgccaggc	agttttaacg	12300
atcagttcgc	cgatgcagat	attcgtaatt	atgtgggcaa	cgtctggtat	cagcgcggaag	12360
tctttatacc	gaaaggttgg	gcagggcagc	aactcgtgct	gcgtttcgat	gcggctcact	12420
attacggcaa	agtgtgggtc	aataatcagg	aatgtatgga	gcatcagggc	ggctatacgc	12480
catttgaagc	cgatgtcacg	ccgtatgtta	ttgccgggaa	aatgtgtacgt	atcacagttt	12540
gtgtgaacaa	cgaactgaac	tggcagacta	tcccgccggg	aatggtgatt	accgcaagaa	12600
acggcaagaa	aaagcagtc	tacttccatg	atcttcttaa	ctacgcgggg	atccatcgca	12660
gcgtaatgct	ctacaccacg	ccgaacacct	gggtggacga	tatcacccgtg	gtgacgcgat	12720
tcgcgcaaga	ctgtaaaccac	gcgtctgttg	actggcaggt	gggtggccaat	gggtggttca	12780
gcgttgaaat	gcgtgatgcg	gatcaacagg	tggttgcaac	tggacaaggc	accagcgagg	12840
ctttgcaagt	ggtgaatccg	caacctctggc	aaccgggtga	aggttatctc	tatgaactgt	12900
acgtcacagc	caaaagccag	acagagtgtg	atatactacc	gctgcgcgtc	ggcatccggt	12960
cagtgggcagt	gaagggcgaa	cagttcctga	tcaaccacaa	accgttctac	tttactgggt	13020
ttggccgtca	tgaagatgcg	gatttgcgcg	gcaaaggatt	cgataacgtg	ctgatgggtg	13080
acgatcacgc	attaatggac	tggattgggg	ccaactccta	ccgtacctcg	cattacctct	13140
acgctgaaga	gatgctcgac	tgggcagatg	aacatggcat	cgtgggtgatt	gatgaaactg	13200
cagctgtcgg	ctttaacctc	tctttaggca	ttggtttcga	agcgggcaac	aagccgaaag	13260
aactgtacag	cgaagaggca	gtcaacgggg	aaactcagca	ggcgcaacta	caggcgatta	13320
aagagctgat	agcgcgtgac	aaaaaccacc	caagcgtggg	gatgtggagt	attgccaacg	13380
aaccggatac	ccgtccgcaa	ggtgcacggg	aatatttctgc	gccactggcg	gaagcaacgc	13440
gtaaactcga	tccgacgcgt	ccgatcacct	gcgtcaatgt	aatgttctgc	gacgctcaca	13500
ccgataccat	cagcgatctc	tttgatgtgc	tgtgcoctgaa	ccgttattac	ggttgggtatg	13560
tccaaagcgg	cgatttgga	acggcagaga	aggtactgga	aaaagaactt	ctggcctggc	13620
aggagaaact	gcatcagccg	attatcatca	ccgaatacgg	cgtggatacg	ttagccgggg	13680
tgcaactaat	gtacaccgac	atgtggagt	aagagtatca	gtgtgcatgg	ctggatatgt	13740
atcacccgct	ctttgatcgc	gtcagcgccg	tgcgtcggtga	acaggtatgg	aatttcgccc	13800
attttgcgac	ctcgcaaggc	atattgcgcg	ttggcggtaa	caagaagggg	atcttcaccc	13860
gcgaccgcaa	accgaagtgc	gcggcttttc	tgctgcaaaa	acgctggact	ggcatgaact	13920
tcggtgaaaa	accgcagcag	ggaggcaaac	aatgaatcaa	caactctcct	ggcgcaacat	13980
ctgcggctac	agcctcggga	attgcgtacc	gagctcgaat	ttccccgatc	gttcaaacat	14040
ttggcaataa	agtttcttaa	gattgaatcc	tgttgcgggt	cttgcgatga	ttatcatata	14100
atttctgttg	aattacgtta	agcatgtaat	aattaacatg	taatgcatga	cgttatttat	14160
gagatgggtt	tttatgatta	gagtcocgca	attatacatt	taatacgcca	tagaaaaaca	14220
aatatagcgc	gcaaaactagg	ataaattatc	gcgcgcggtg	tcacttatgt	tactagatcg	14280
ggaattcgat	atcaagcttg	gcactggccg	tgcgtttaca	acgtcgtgac	tgggaaacc	14340
ctggcgcttac	ccaacttaat	cgccttgcag	cacatcccc	tttcgcccagc	tggcgtaata	14400
cggaagaggc	cgcacccgat	cgccttccc	acagtttgcg	cagcctgaat	ggcgaatgct	14460
agagcagctt	gagcttggat	cagattgtcg	tttcccgcc	tcagtttaaa	ctatcagtgt	14520

ttgacaggat	atattggcgg	gtaaacctaa	gagaaaagag	cgtttattag	aataacggat	14580
attttaaagg	gcgtgaaaag	gtttatccgt	tcgtccattt	gtatgtg		14627

<210> 110  
 <211> 9080  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> p18attBZeo(6XHS4)2eGFP Plasmid

<400> 110						
cagttgcccgg	ccggggtcgcg	cagggcggaac	tcccgcctccc	acgggtgctc	gccgatctcg	60
gtcatggcccg	gcccggaggcg	gtcccgggaag	ttcgtgggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggcccgcg	caccacacacc	caggccagggg	tggtgtccgg	caccacctgg	180
tcctggaccg	cgctgatgaa	cagggtcacg	tcgtcccggg	ccacaccggc	gaagtctgct	240
tccacgaagt	cccgggagaa	cccggagccgg	tcggtccaga	actcgaccgc	tcgggcgacg	300
tcgcgccggg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgatata	gaattcctgc	420
agcccccgcg	atccgctcac	ggggacagcc	cccccccaaa	gccccagggg	atgtaattac	480
gtccctcccc	cgctaggggg	cagcagcgag	ccgccccggg	ctccgctccg	gtccggcgct	540
cccccccgcat	ccccgagccg	gcagcggtcg	gggacagccc	gggcacgggg	aagggtggcac	600
gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	cctgggggat	660
acggggccgc	ggatccgctc	acggggacag	ccccccccca	aagccccag	ggatgtaatt	720
acgtccctcc	cccgttaggg	ggcagcagcg	agccgcctcg	ggctccgctc	cggtccggcg	780
ctcccccgcg	atccccgagc	cggcagcggt	cggggacagc	ccggggcacgg	ggaaggtggc	840
acgggatcg	tttctctgga	acgcttctcg	ctgctctttg	agcctgcaga	cacctggggg	900
atacggggcc	gcggatccgc	tcacggggag	agcccccccc	caaagcccc	agggatgtaa	960
ttacgtccct	cccccgctag	ggggcagcag	cgagcccgcc	gggggtccgc	tcgggtccgg	1020
cgctcccccc	gcatccccga	gccggcagcg	tgccggggaca	gcccggggcac	ggggaaggtg	1080
gcacggggate	gctttccctc	gaacgcttct	cgctgctctt	tgagcctgca	gacacctggg	1140
ggatacgggg	ccgcggatcc	gctcacgggg	acagcccccc	cccaaagccc	ccagggatgt	1200
aattacgtcc	ctcccccgct	aggggggcag	agcgagccgc	ccgggggtcc	gctccgggtc	1260
ggcgctcccc	ccgcattccc	gagccggcag	cgctgcgggg	cagccccggg	acgggggaag	1320
tggcacgggga	tcgctttcct	ctgaacgctt	ctcgctgctc	tttgagcctg	cagacacctg	1380
ggggatacgg	ggcccgcgat	ccgctcacgg	ggacagcccc	cccccaaagc	ccccagggat	1440
gtaattacgt	ccctcccccg	ctagggggga	gcagcgagcc	gcccgggggt	ccgctccggg	1500
ccggcgctcc	ccccgcattc	ccgagccggc	agcgtgcggg	gacagccccg	gcacggggaa	1560
gggtggcacgg	gatcgcttcc	ctctgaacgc	ttctcgctgc	tccttgagcc	tcgacagacc	1620
tgggggatac	ggggcccgcg	atccgctcac	ggggacagcc	ccccccccaa	gccccagggg	1680
atgtaattac	gtccctcccc	cgctaggggg	cagcagcgag	ccgccccggg	ctccgctccg	1740
tgccagtcgc	cccccgcgct	cccccgagcc	gcagcgtgcg	gggacagccc	gggcacgggg	1800
aagggtggcac	gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttattaata	gtaatcaatt	acggggctat	1920
tagttcatag	cccatatag	gagttccgct	ttacataact	tacggtaaat	ggcccgcttg	1980
gctgaccgcc	caacgacccc	cgcccattga	cgctcaataat	gacgtatggt	cccatagtaa	2040
cgccaatagg	gaactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgcccact	2100
tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagtaac	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcata	gctattacca	tgggtcgagg	tgagccccac	gttctgcttc	2280
actctcccca	tctccccccc	ctccccaccc	ccaattttgt	atttatttat	tttttaatta	2340
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cggggcgggg	2400
cgaggggcg	ggcgggggcg	ggcgggagag	tgccggcgga	gccaatcaga	gcggcgcgct	2460
ccgaaagt	ccttttatgg	cgaggcgggc	gcggcgggcg	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtgcgtg	cggtgccttc	gccccgtgcc	ccgctcccg	ccgcctcgcg	2580
ccgccccgg	cggtctgac	tgaccgcgtt	actccacag	gtgagcgggc	gggacggccc	2640
ttctctccg	ggctgtaatt	agcgcttggt	ttaatgacgg	ctcgtttctt	ttctgtgggt	2700
gcgtgaaagc	cctaaagggc	tcggggaggg	ccctttgtgc	gggggggagc	ggctcggggg	2760
gtgcgtcgct	gtgtgtgtgc	gtggggagcg	cccgctgctg	cccgcgctgc	ccggcggtcg	2820
tgagcgctgc	gggcggcgcg	cggggctttg	tgccgtccgc	gtgtgcgcga	ggggcagcgc	2880
gcccggggcg	gtgcccccg	gtgcgggggg	gctgcgaggg	gaacaaaggc	tgctgctggg	2940
gtgtgtcgct	gggggggtga	gcaggggggt	tgggcgcggc	ggtcgggctg	taaccccccc	3000
ctgcaccccc	ctccccgag	tgctgagcac	ggccccgctt	cggggtgcgg	gctccgtgcg	3060
gggcgtggcg	cggggctcgc	cggtccgggc	gggggggtgg	ggcagggtgg	gggtccgggc	3120
ggggcgggcg	cgcctcgggc	cgggggaggc	tcggggggag	ggcgcgcgcg	ccccggagcg	3180
ccggcggtcg	tcgagggcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
gggcgcaggg	acttcccttg	tcaccaatct	ggcgagaccg	aaatctggga	ggcgcccgcc	3300



cacccccctct	agcggggcgcg	ggcgaagcgcg	tgcgggcgcgc	gcaggaagga	aatggggcggg	3360
gagggccttc	gtgcgtcgcc	gcgcgcgcgt	ccccctctcc	atctccagcc	tcgggggctgc	3420
cgcaggggga	cggctgcctt	cggggggggac	ggggcagggc	gggggttcggc	ttctggcgctg	3480
tgaccggcg	ctctagagcc	tctgctaacc	atgttcatgc	cttcttcttt	ttcctacagc	3540
tcctgggcaa	cgtgctgggt	gttggtgctgt	ctcatcattt	tggcaaaagaa	ttcgccacca	3600
tggtgagcaa	gggcgaggag	ctgttcaccg	gggtgggtgcc	catcctggtc	gagctggagc	3660
gcgacgtaaa	cggccacaag	ttcagcgtgt	ccggcgaggg	cgagggcgat	gccacctacg	3720
gcaagctgac	cctgaagttc	atctgcacca	ccggcaagct	gcccgtgcc	tgccccaccc	3780
tcgtgaccac	cctgacctac	ggcgtgcagt	gcttcagccg	ctaccccgac	cacatgaagc	3840
agcacgactt	cttcaagtcc	gccatgcccc	aaggctacgt	ccaggagcgc	accatcttct	3900
tcaaggacga	cggcaactac	aagaccgcg	ccgaggtgaa	gttcgagggc	gacacctgg	3960
tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	cggcaacatc	ctgggggcaca	4020
agctggagta	caactacaac	agccacaacg	tctatatcat	ggccgacaag	cagaagaacg	4080
gcatcaaggt	gaacttcaag	atccgccaca	acatcgagga	cggcagcggtg	cagctcgccg	4140
accactacca	gcagaacacc	cccatcggcg	acggccccgt	gctgctgccc	gacaaccact	4200
acctgagcac	ccagtccgcc	ctgagcaaaag	accccaacga	gaagcgcgat	cacatggtcc	4260
tgctggagtt	cgtgaccgcc	gccgggatgc	ctctcggcac	ggacgagctg	tacaagtaag	4320
aattcactcc	tcaggtgcag	gctgcctatc	agaaggtggg	ggctgggtgtg	gccaatgccc	4380
tggtcacaaa	ataccactga	gatctttttc	cctctgccaa	aaattatggg	gacatcatga	4440
agccccttga	gcactgtgact	tctggctaata	aaaggaaatt	tattttcatt	gcaactgtgt	4500
gttggaaatt	tttgtgtctc	tcactcgga	ggacatatgg	gagggcaaat	catttaaaac	4560
atcagaatga	gtatttgggt	tagagtttgg	caacatatgc	catatgctgg	ctgccatgaa	4620
caaagggtggc	tataaagagg	tcatacgtat	atgaacacgc	ccccgtctgt	ccattcctta	4680
ttccatagaa	aagccttgac	ttgaggttag	atttttttta	tattttgttt	tgtgttattt	4740
ttttcttttaa	catccctaaa	attttccctta	catgttttac	tagccagatt	tttctcctc	4800
tcctgactac	tccagtcctc	agctgtccct	cttctcttat	gaagatccct	cgacctgcag	4860
cccaagcttg	catgctgca	ggtcgactct	agtggatccc	ccgccccgta	ttccccaggt	4920
gtctgcaggg	tcaaagagca	gcgagaagcg	ttcagaggaa	agcgatcccc	tgccaccttc	4980
cccgtgcccc	ggctgtcccc	gcacgctgcc	ggctcgggga	tgccgggggga	gcgcgcggacc	5040
ggagcgggag	cccggggcg	tcgctgtgct	ccccatagcg	gggaggggacg	taattacatc	5100
cctggggggt	ttgggggggg	gctgtcccc	tgagcggatc	cgcggccccg	tatccccacg	5160
gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	aaagcgatcc	cgtgccacct	5220
tcgccgtgcc	cgggctgtcc	ccgcacgctg	ccggctcggg	gatgcggggg	gagcgccgga	5280
ccggagcggg	gccccggggc	gctcgctgct	gccccctagc	gggggagggg	cgtaattaca	5340
tccttggggg	ctttgggggg	gggctgtccc	cgtgagcgga	tcgcgggccc	cgtatcccc	5400
aggtgtctgc	aggctcaaag	agcagcgaga	agcgttcaga	ggaaagcgat	cccgtgccac	5460
cttccccgtg	cccgggctgt	cccgcacgc	tgccgggtcg	gggatgcggg	gggagcgccg	5520
gaccggagcg	gagccccggg	cggtctgctg	ctgcccccta	gcgggggagg	gacgtaatta	5580
catccctggg	ggctttgggg	gggggctgtc	gtcgtagcg	gatccggcg	cccgtatccc	5640
ccaggtgtct	gcaggctcaa	agagcagcga	gaagcgttca	gaggaaagcg	atcccgtgcc	5700
accttccccg	tgccccgggt	gtccccgcac	gctgcgggct	cggggatgcg	gggggagcgc	5760
cggaccggag	cggagccccg	ggcggtctgc	tgctgcccc	tagcggggga	gggacgtaat	5820
tacatccctg	ggggctttgg	gggggggctg	tcctcgtag	cggatccgcg	gccccgtatc	5880
ccccaggtgt	ctgcaggctc	aaagagcagc	gagaagcgtt	cagaggaaaag	cgatccccgt	5940
ccaccttccc	cgtgcccggg	ctgtccccgc	acgctgcccc	ctcggggatg	cgggggggagc	6000
gccggaccgg	agcggagccc	cgggcggtct	gctgctgccc	cctagcgggg	gaggggacgta	6060
attacatccc	tgggggcttt	gggggggggg	tgtccccgtg	agcggatccg	cggccccgta	6120
tcctccaggt	gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	agcgatcccc	6180
tgccaccttc	cccgtgcccc	ggctgtcccc	gcacgctgcc	ggctcgggga	tgccgggggga	6240
gcgcgggacc	ggagcgggag	cccggggcgg	tcgctgtctg	ccccatagcg	gggaggggacg	6300
taattacatc	cctggggggt	ttgggggggg	gctgtcccc	tgagcggatc	cgcgggggctg	6360
caggaattcg	taatcatggt	catagctgtt	tcctgtgtga	aattgttatc	cgctcacaat	6420
tcacacaaac	atacagagcg	gaagcataaa	gtgtaaagcc	tgggggtgct	aatgagttag	6480
ctaactcaca	ttaattgcgt	tgcgctcact	gcccgttttc	cagtccggga	acctgtcggtg	6540
ccagctgcat	taatgaatcg	gccaacgcgc	ggggagaggc	ggtttgcgta	ttggggcgctc	6600
ttcgcttccc	tcgctcactg	actcgctgcg	ctcggtcggt	cggctgcggc	gagcgggtatc	6660
agctcactca	aaggcggtaa	tacggttatc	cacagaatca	ggggataaacg	caggaaagaa	6720
catgtgagca	aaaggccagc	aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggtt	6780
tttccatagg	ctccgcccc	ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagaggtg	6840
gcgaaaccgc	acaggactat	aaagatacca	ggcgtttccc	cctggaaagct	ccctcggtgc	6900
ctctcctgtt	ccgacctgct	cgcttaccgg	atacctgtcc	gcctttctcc	cttcggggaag	6960
cgtggcgctt	tctcatagct	cacgctgtag	gtatctcagt	tcggtgtagg	tcgttcgctc	7020
caagctgggc	tggtgtgcag	aacccccgt	tcagcccgac	cgctgcgcct	tatccggtaa	7080
ctatcgctct	gagtccaacc	cggtaagaca	cgacttatcg	ccactggcag	cagccactgg	7140
taacaggatt	agcagagcga	ggtatgtagg	cgggtgctaca	gagttcttga	agtgggtggc	7200
taactacggc	tacactagaa	ggacagtatt	tggtatctgc	gctctgctga	agccagttac	7260
cttcggaaaa	agagttggta	gctcttgatc	cggcaacaa	accaccgctg	gtagcgggtg	7320

tttttttgggt	tgcaagcagc	agattacgcg	cagaaaaaaa	ggatctcaag	aagatccttt	7380
gatcttttct	acgggggtctg	acgctcagtg	gaacgaaaaac	tcacgttaag	ggatttttgg	7440
catgagatta	tcaaaaaggga	tcttcaccta	gatccttttta	aattaaaaaat	gaagtttttaa	7500
atcaatctaa	agtatatatg	agtaaacttg	gtctgacagt	taccaatgct	taatcagtg	7560
ggcacctatc	tcagcgatct	gtctatctcg	ttcatccata	ggtgcctgac	tccccgctcg	7620
gtagataaact	acgatacggg	agggccttacc	atctggcccc	agtgcctgcaa	tgataccgcg	7680
agacccacgc	tcaccggctc	cagattttatc	agcaataaac	cagccagccg	gaagggccga	7740
gcccagaagt	ggctctgcaa	ctttatccgc	ctccatccag	tctattaatt	ggtgcgggga	7800
agctagagta	agtagttcgc	cagttaatag	tttgccgaac	ggtgttgcca	ttgctacagg	7860
catcgtgggtg	tcacgctcgt	cgttttggtat	ggcttcattc	agctccggtt	cccaacgatc	7920
aaggcgagtt	acatgatccc	ccatgttggtg	caaaaaagcg	gtagctcct	tcggtcctcc	7980
gatcgtttgtc	agaagtaagt	tggccgcagt	ggtatcactc	atgggttatgg	cagcactgca	8040
taattctctt	actgtcaatg	catccgttaag	atgcttttct	gtgactggtg	agtactcaac	8100
caagtctagt	tgagaatagt	gtatgcgggtg	accgagttgc	tcttgcccgg	cgtcaataacg	8160
ggataataacc	gcgccacata	gcagaacttt	aaaagtgtctc	atcattggaa	aacgttcttc	8220
ggggcgaaaa	ctctcaagga	tcttaccgct	ggtgagatcc	agttcgatgt	aaccactcgc	8280
tgaccccaac	tgatcttcag	catcttttac	ttccaccagc	gtttctgggt	gagcaaaaaac	8340
aggaaggcaa	aatgccgcaa	aaaagggaat	aaggcgacac	cggaaatggt	gaatactcat	8400
actcttcctt	tttcaatatt	attgaagcat	ttatcagggg	tattgtctca	tgagcggata	8460
catatttgaa	tgtattttag	aaaataaaca	aatagggggt	ccgcgcacat	ttccccgaaa	8520
agtgccacct	gacgtagtta	acaaaaaaaa	gcccgcgcga	gcgggcttta	ttaccaagcg	8580
aagcgccatt	cgccattcag	gctgcgcaac	tgttgggaag	ggcgatcggg	gcggggcctct	8640
tcgctattac	gccagctggc	gaaaggggga	tgtgctgcaa	ggcgattaa	ttgggtaacg	8700
ccagggtttt	cccagtcacg	acgttgtaaa	acgacggcca	gtccgtaata	cgactcactt	8760
aaggccttga	ctagagggtc	gacggtatac	agacatgata	agatacattg	atgagtttgg	8820
acaaaccaca	actagaatgc	agtgaaaaaa	tgtccttatt	tgtgaaattt	gtgatgctat	8880
tgcttttatt	gtaaccatta	taagctgcaa	taaacaagtt	ggggtggg	aagaactcca	8940
gcatgagatc	cccgcgctgg	aggatcatcc	agccggcgct	ccggaaaacg	attccgaagc	9000
ccaacctttc	atagaaggcg	gcggtggaat	cgaaatctcg	tagcacgtgt	cagtcctgct	9060
cctcggccac	gaagtgcacg					9080

<210> 111  
 <211> 4223  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pLIT38attBBSRpolyA10 Plasmid

<400> 111						
gttaactacg	tcagggtggca	cttttcgggg	aaatgtgcgc	ggaaccctta	tttgtttatt	60
tttctaaata	cattcaataa	tgtatccgct	catgagacaa	taaccctgat	aaatgcttca	120
ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	ttattccctt	180
ttttgcggca	ttttgccttc	ctgtttttgc	tcaccagaaa	acgctgggtga	aagtataaaga	240
tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	acagcggtaa	300
gacctttgag	agttttcgcc	ccgaagaacg	ttctccaatg	atgagcactt	ttaaagttct	360
gctatgtggc	gcggtattat	cccggtgttg	cgccggggcaa	gagcaactcg	gtcgcgcgat	420
acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	atcttacgga	480
tgccatgaca	gtaagagaat	tatgcagtcg	tgccataacc	atgagtata	acactgcggc	540
caacttactt	ctgacaacga	tccgaggacc	gaaggagcta	accgcttttt	tgacacaact	600
gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	ccataccaaa	660
cgacgagcgt	gacaccacga	tgccgttagc	aatggcaaca	acgttgcgca	aactattaac	720
tgccgaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	aggcggataa	780
agttgcagga	ccacttctgc	gctcggccct	tccggctggc	tggtttattg	ctgataaaatc	840
tgagcgggt	gagcgtgggt	ctcggcggtat	cattgcagca	ctggggccag	atggtaagcc	900
ctcccgtatc	gtagttatct	acacgacggg	gagtcaggca	actatggatg	aacgaaatag	960
acagatcgct	gagatagggt	cctcactgat	taagcattgg	taactgtcag	accaagttta	1020
ctcatatata	ctttagattg	atttaccccg	gttgataatc	agaaaagccc	caaaaacagg	1080
aagattgtat	aagcaaatat	ttaaattgta	aacgttaata	ttttgttaaa	attcgcgtta	1140
aatttttggt	aaatcagctc	attttttaac	caataggccg	aaatcgccaa	aatcccttat	1200
aaatcaaaag	aatagcccca	gatagggttg	agtgttgttc	cagtttggaa	caagagtcca	1260
ctattaaaga	acgtggactc	caacgtcaaa	ggggcgaaaa	ccgtctatca	gggcgatggc	1320
ccactacgtg	aacctacacc	caaatcaagt	tttttggggt	cgaggtgccc	taagacacta	1380
aatcggaacc	ctaaagggag	cccccgattt	agagcttgac	ggggaaagcg	aacgtggcga	1440
gaaaggaagg	gaagaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcgggtca	1500
cgtgcgcgt	aaccaccaca	ccgcgcgcgc	ttaatgcgcc	gctacagggc	gcgtaaaagc	1560
atctaggtga	agatcctttt	tgataatctc	atgacaaaaa	tcccttaacg	tgagttttcg	1620



ttccactgag	cgtagacccc	cgtagaaaa	atcaaaggat	cttcttgaga	tccttttttt	1680
ctgcgcgtaa	tctgctgctt	gcaaacaaaa	aaaccaccgc	taccagcggg	ggtttggttg	1740
ccggatcaag	agctaccaac	tctttttccg	aaggtaactg	gcttcagcag	agcgagata	1800
ccaaatactg	ttcttctagt	gtagccgtag	ttaggccacc	acttcaagaa	ctctgttagca	1860
ccgcctacat	acctcgctct	gctaactcctg	ttaccagtg	ctgctgccag	tggcgataag	1920
tcgtgtctta	ccgggttgga	ctcaagacga	tagttaccgg	ataaggcgca	gcggtcgggc	1980
tgaacggggg	gttcgtgcac	acagcccagc	ttggagcgaa	cgacctacac	cgaactgaga	2040
tacctacagc	gtgagctatg	agaaagcgcc	acgcttcccg	aagggagaaa	ggcggacagg	2100
tatccggtaa	gcggcagggg	cggaacagga	gagcgacga	gggagcttcc	agggggaaac	2160
gcctgggtac	tttatagtc	tgctcggttt	cgccacctct	gacttgagcg	tcgatttttg	2220
tgatgtctgt	caggggggag	gagcctatgg	aaaaacgcca	gcaacgcgcc	ctttttacgg	2280
ttcctggcct	tttgctggcc	ttttgctcac	atgtaatgtg	agtttagctca	ctcattaggc	2340
accccagggt	ttacacttta	tgcttccggc	tcgtatgttg	tgtggaattg	tgagcggata	2400
acaatttcac	acaggaacaa	gctatgacca	tgattacgcc	aagctacgta	atagcactca	2460
ctagtggggc	ccgtgcaatt	gaagccggct	ggcgccaagc	ttctctgcag	gattgaagcc	2520
tgctttttta	tactaacttg	agcgaaatct	ggatcaccat	gaaaacattt	aacattttct	2580
aacaagatct	agaattagta	gaagtagcga	cagagaagat	tacaatgctt	tatgaggata	2640
ataaacatca	tgtgggagcg	gcaattcgta	cgaaaacagg	agaaatcatt	tcggcagtac	2700
atatgaagc	gtatatagga	cgagtaactc	tttgtgcaga	agccattgcy	attggtagtg	2760
cagtttcgaa	tggaacaaa	gattttgaca	cgattgtagc	tgtagacac	ccttattctg	2820
acgaagtaga	tagaagtatt	cgagtggtaa	gtccttgtgg	tatgtgtagg	gagttgattt	2880
cagactatgc	accagattgt	tttgtgttaa	tagaaatgaa	tggcaagtta	gtcaaaacta	2940
cgattgaaga	actcattcca	ctcaaatata	cccgaatata	aaagttttac	cataccaagc	3000
ttggctgctg	cctgaggctg	gacgacctcg	cgagattcta	ccggcagtg	aaatccgtcg	3060
gcattccagg	aaccagcagc	ggctatccgc	gcattccatg	ccccgaactg	caggagtggg	3120
gaggcacgat	ggcgcgtttg	gtccggatct	ttgtgaagg	accttacttc	tgtggtgtga	3180
cataattgga	caactacct	acagagattt	aaagctctaa	ggtaaatata	aaatttttaa	3240
gtgtataatg	tggttaaact	ctgattctaa	ttgtttgtgt	atttttagatt	ccaacctatg	3300
gaactgatga	atgggagcag	tggtggaatg	cctttaatga	ggaaaacctg	ttttgctcag	3360
aagaaatgcc	atctagtgat	gatgaggcta	ctgctgactc	tcaacattct	actcctccaa	3420
aaaagaagag	aaaggtagaa	gaccccaagg	actttccttc	agaattgcta	agttttttga	3480
gtcattgctg	gtttcagtaa	gtttcctttg	ccttgctttg	tattttacacc	acaaaggaaa	3540
aagctgcact	gctatacaag	aaaatttatgg	aaaaatatct	tgtaaccttt	ataagtaggc	3600
ataacagtta	taatcataaac	atactgtttt	ttcttactcc	acacaggcat	agagtgtctg	3660
ctattaataa	ctatgctcaa	aaatttgtgta	ccttttagctt	tttaatttgt	aaaggggtta	3720
ataaggaata	tttgatgtat	agtgccttga	ctagagatca	taatcagcca	taccacattt	3780
gtagaggttt	tacttgcttt	aaaaaacctc	ccacacctcc	ccctgaacct	gaaacataaa	3840
atgaatgcaa	ttgttttgt	taacttggtt	ataatggctt	ataatgggta	caaataaagc	3900
aatagcatca	caaatttcac	aaataaagat	ccacgaattc	gctagcttcg	gccgtgacgc	3960
gtctccggat	gtacaggcat	gcgtcgaccc	tctagtcaag	gccttaagt	agtcgtatta	4020
cggactggcc	gtcgttttac	aacgtcgtga	ctgggaaaac	cctggcggtta	cccaacttaa	4080
tcgccttgca	gcacatcccc	ctttcgccag	ctggcgtaat	agcgaagagg	cccgccacca	4140
tcgcccctcc	caacagttgc	gcagcctgaa	tggcgaatgg	cgcttcgctt	ggtaataaag	4200
cccgcttcgg	cgggcttttt	ttt				4223

<210> 112

<211> 5855

<212> DNA

<213> Artificial Sequence

<220>

<223> pCX-LamIntR Plasmid

<400> 112

gtcgacattg	attattgact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccataat	ggagttccgc	gttacataac	ttacggtaaa	tggcccgcc	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtac	240
atcaagtgt	tcatatgcc	agtacgcccc	ctattgacgt	caatgacggg	aaatggcccg	300
cctggcatta	tgcccagtac	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tatttagtcat	cgctattacc	atgggtcgag	gtgagcccca	cgctctgctt	cactctcccc	420
atctcccccc	cctccccacc	cccaattttg	tattttattt	tttttttaatt	atttttgtgca	480
gcgatggggg	cggggggggg	ggggggcgcg	gcaggcggg	gcggggcgcg	gcgagggggc	540
gggcgggggg	aggcggagag	gtgcggcggc	agccaatcag	agcggcgcg	tccgaaagtt	600
tccttttatg	gcgaggcggc	ggcggcgggc	cccctataaa	aagcgaagcg	gcggcggggc	660
gggagtcgct	gcgttgccct	cgcccgctgc	cccctccgc	gccgcctcgc	gccggccgcc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtgagcggg	cgggacggcc	cttctcctcc	780

gggctgtaat	tagcgcttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgcgtgaaa	840
ccttaaaggg	ctccgggagg	gcoctttgtg	cgggggggag	cggtcgggg	ggtgctgctg	900
tgtgtgtgtg	cgtggggagc	gocgcgtgcg	gcccgcgctg	cccggcggtc	gtgagcgctg	960
cgggcgcgcg	gcggggcttt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggccgggggc	1020
ggtgccccgc	ggtgccccgc	ggtgccccgc	ggaacaaagg	ctgcgtgctg	ggtgtgtgctg	1080
tgggggggtg	agcagggggg	gtggggcgcg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cgcccggtc	tcgggtgctg	ggctcgtg	ggggcggtgg	1200
gcggggctcg	ccgtgccccg	cggggggtgg	cggcaggtgg	gggtgccccg	cgggcggggg	1260
ccgcctcggg	ccggggaggg	ctcgggggag	gggcgcggcg	gccccgggag	gcccggcggt	1320
gtcgagggcg	ggcgagccgc	agccattgcc	ttttatggta	atcggtgcgag	agggcgcgagg	1380
gacttccttt	gtcccaaatc	tggcgaggcc	gaaatctggg	aggcgccgcc	gcacccccctc	1440
tagcgggcg	gggcgaagcg	gtgcggcgcc	ggcaggaagg	aaatgggcgg	ggagggcctt	1500
cgtgcgtcgc	cgcgcgcgcg	tcccttctc	cactccagc	ctcggggctg	ccgcaggggg	1560
acggctgcct	tcggggggga	cggggcaggg	cggggttcgg	cttctggcgt	gtgaccggcg	1620
gctctagagc	ctctgctaac	catgttccat	ccttcttctt	tttccctacag	ctcctgggca	1680
acgtgctggt	tgttgtgctg	tctcatcatt	ttggcaaaag	attcatggga	agaagcgcaa	1740
gtcatgagcg	ccgggattta	ccccctaacc	tttatataag	aaacaatgga	tattactgct	1800
acagggaccc	aaggacgggt	aaagagtttg	gattaggcag	agacaggcga	atcggaatca	1860
ctgaagctat	acaggccaac	attgagttat	tttcaggaca	caaacacaag	cctctgacag	1920
cgagaatcaa	cagtgataat	tccgttacgt	tacattccatg	gcttgatcgc	tacgaaaaaa	1980
tcctggccag	cagaggaatc	aagcagaaga	cactcataaa	ttacatgagc	aaaattaaag	2040
caataaggag	gggtctgcct	gatgctccac	ttgaagacat	caccacaaaa	gaaattgcgg	2100
caatgctcaa	tggatacata	gacgagggca	aggcggcgtc	agccaagtta	atcagatcaa	2160
cactgagcga	tgcattccga	gaggcaatag	ctgaaggcca	tataacaaca	aaccatgtcg	2220
ctgccactcg	cgcagcaaaa	tctagagtaa	ggagatcaag	acttacggct	gacgaatacc	2280
tgaaaattta	tcaagcagca	gaatcatcac	catgttggct	cagacttgca	atggaaactgg	2340
ctgttggttac	cgggcaacga	gttgggtgatt	tatgcgaaat	gaagtggctc	gatatacgtag	2400
atggatatct	ttatgtcgag	caaagcaaaa	caggcgtaaa	aattggccatc	ccaacagcat	2460
tgcataattga	tgetctcgga	atatcaatga	aggaacacat	tgataaatgc	aaagagattc	2520
ttggcgggaga	aaccataatt	gcattctactc	gtcgcgaaac	gctttccatcc	ggcacagtat	2580
caaggtatct	tatgcgcgca	cgaaaagcat	caggctcttc	cttcgaaggg	gatccgccta	2640
cctttcacga	gttgcgcagt	ttgtctgcaa	gactctatga	gaagcagata	agcgataagt	2700
ttgctcaaca	tcttctcggg	cataagtcgg	acaccatggc	atcacagtat	cgtgatgaca	2760
gaggcagggga	gtgggacaaa	attgaaatca	aataagaatt	cactcctcag	gtgcaggctg	2820
cctatcagaa	ggtggtggct	ggtgtggcca	atgccctggc	tcacaaatac	cactgagatc	2880
tttttccctc	tgcacaaaat	tatggggaca	tcatgaagcc	ccttgagcat	ctgacttctg	2940
gctaataaag	gaaattttatt	ttcattgcaa	tagtgtgttg	gaattttttg	tgtctctcac	3000
tcggaaggac	atatgggagg	gcaaatcatt	taaaacatca	gaatgagtat	ttggtttaga	3060
gtttggcaac	atatgccata	tgtgtgctgc	catgaacaaa	ggtggctata	aagaggctcat	3120
cagtatatga	aacagccccc	tgtgtgctat	tccttatctc	atagaaaagc	ccttgacttga	3180
gggtagatct	ttttttagatt	ttgttttggtg	ttattttttt	ctttaacatc	cctaaatttt	3240
tccttacatg	ttttactagc	cagatttttct	ctcctctcct	gactactccc	agtcatagct	3300
gtccctcttc	tcttatgaag	atccctcgac	ctgcagccca	agcttggcgt	aatcatggct	3360
atagctgttt	cctgtgtgaa	attgttatcc	gctcacaaat	ccacacaaca	tacgagccgg	3420
aagcataaag	tgtaaagcct	ggggtgccta	atgagtgagc	taactcacat	taattgcgtt	3480
gcgcctcactg	cccgctttcc	agtcgggaaa	cctgtcgtgc	cagcggatcc	gcattctcaat	3540
tagtcagcaa	ccatagtcct	gcccctaact	ccgcccattc	cgcccctaac	tcgcgccagt	3600
tcgcgccatt	ctccgccccca	tggctgacta	attttttttt	tttatgcaga	ggccgaggcc	3660
gcctcggect	ctgagctatt	ccagaagtag	tgaggaggct	tttttgagg	cctaggcttt	3720
tgcacaaaag	taacttgggt	attgcagctt	ataatgggta	caaataaagc	aatagcatca	3780
caaatttcac	aaataaagca	ttttttttcac	tgcattctag	ttgtgggttg	tcacaaactca	3840
tcaatgtatc	ttatcatgtc	tggatccgct	gcattaatga	atcgcccaac	gcgcggggag	3900
aggcggtttg	cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	3960
cgttcggctg	cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	4020
atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaagg	cagcaaaaagg	ccaggaaccg	4080
taaaaaggcc	gcgttgctgg	cgttttttcca	taggctccgc	ccccctgacg	agcatcaca	4140
aaatcgacgc	tcaagtccga	ggtggcgaaa	ccgcacagga	ctataaagat	accaggcggt	4200
tccccctgga	agctccctcg	tgcgtctctc	tggtccgacc	ctgcccgtta	ccggatatct	4260
gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcaa	tgctcacgct	gtaggtatct	4320
cagttcggtg	taggtcgttc	gctccaagct	gggtgtgtgtg	cacgaacccc	ccgttcagcc	4380
cgaccgctgc	gccttatccg	gtaactatcg	tcttgagtc	aacccggtaa	gacacgactt	4440
atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	taggcgggtc	4500
tacagagttc	ttgaagtggt	ggcctaacta	cggtacact	agaaggacag	tatttggtat	4560
ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	4620
acaaaccacc	gctggtagcg	gtgggttttt	tggtttgcaag	cagcagatta	cgcgacagaa	4680
aaaaggactc	caagaagatc	ctttgatctt	tctacgggg	tctgacgctc	agtggaacga	4740
aaactcacgt	taagggattt	tggctcatgag	attatcaaaa	aggatcttca	cctagatcct	4800

tttaaattaa	aaatgaagtt	ttaaatacaat	ctaaagtata	tatgagtaaa	cttgggtctga	4860
cagttaccaaa	tgcttaaatca	gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcatc	4920
catagttgcc	tgactccccg	tcgtgtagat	aactacgata	cgggagggct	taccatctgg	4980
ccccagtgt	gcaatgatac	cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	5040
aaaccagcca	gccggaaggg	ccgagcgcag	aagtggctct	gcaactttat	ccgcctccat	5100
ccagttctatt	aattgtttgcc	gggaagctag	agtaagtagt	tcgcccagtt	atagtttgcg	5160
caacgttggt	gccattgcta	caggcatcgt	gggtgcacgc	tcgtcgtttg	gtatggcttc	5220
attcagctcc	ggttcccaac	gatcaaggcg	agttacatga	tcccccatgt	tgtgcaaaaa	5280
agcgggttagc	tccttcgggtc	ctccgatcgt	tgtcagaagt	aagttggcgg	cagtggttatc	5340
actcatgggt	atggcagcac	tgcataattc	tcttactgtc	atgccatccg	taagatgctt	5400
ttctgtgact	ggtagtgact	caaccaagtc	attctgagaa	tagtgatatgc	ggcgaccgag	5460
ttgtctcttg	ccggcggtcaa	tacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	5520
gtcatcatt	ggaaaaagtt	cttcggggcg	aaaactctca	aggatcttac	cgctgttgag	5580
atccagttcg	atgtaaccca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	5640
cagcgtttct	gggtgagcaa	aaacaggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	5700
gacacggaaa	tggtgaatac	tcatactctt	ctttttcaa	tattattgaa	gcattatca	5760
gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	5820
ggttccgcgc	acatttcccc	gaaaagtgcc	acctg			5855

<210> 113

<211> 4346

<212> DNA

<213> Artificial Sequence

<220>

<223> pSV40-193AttpsensePur Plasmid

<400> 113

ccggtgccc	caccatcccc	tgaccacgc	ccctgacccc	tcacaaggag	acgaccttcc	60
atgaccgagt	acaagcccac	ggtgcgcctc	gccacccgcg	acgacgtccc	ccgggcccgt	120
cgcacccctcg	ccgcgcgcgtt	cgccgactac	cccgccacgc	gccacaccgt	cgacccggac	180
cgccacatcg	agcgggtcac	cgagctgcaa	gaactcttcc	tcacgcgcgt	cgggctcgac	240
atcggcaagg	tgtgggtcgc	ggacgacggc	gcgcgcggtg	cggctctggac	cacgcgggag	300
agcgtcgaag	cgggggcggt	gttcgcccag	atcggcccgc	gcattggccga	ggtgagcggg	360
tcccggctgg	ccgcgcagca	acagatggaa	ggcctcctgg	cgccgcacccg	gcccaggag	420
cccgcggtgg	tcctggccac	cgtcggcgctc	tcgcccgaac	accaggggcaa	gggtctgggc	480
agcgcgcgtcg	tgctcccccg	agtggaggcg	gccgagcgcg	ccgggggtgcc	cgccctcctg	540
gagacctccg	cgccccgcaa	cctccccctc	tacgagcggc	tcggcttcac	cgccacccgc	600
gacgtcgagg	tgcccgaagg	accgcgcacc	tgggtgcatga	cccgcgaagcc	cgggtgcctga	660
cgccccgc	acgaccgcga	gcgcgcgacc	gaaaggagcg	cacgacccca	tggtctccgac	720
cgaagcccgac	ccgggcggcc	ccgcgcgacc	cgcacccgcc	cccaggggcc	accgactcta	780
gaggatcata	atcagccata	ccacatttgt	agagggttta	cttgcttttaa	aaaacctccc	840
acacctcccc	ctgaacctga	aacataaaat	gaattgcaatt	gttggttgta	acttggttat	900
tcagctttat	aattggtata	aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	960
tttttcaactg	cattgtagtt	gtgggttgctc	caaactcatc	aattgtatctt	atcatgtctg	1020
gatccgcgc	ggatccttaa	ttaagtctag	agtcgactgt	ttaaaacctgc	aggcatgcaa	1080
ccttggcgta	atcatgggtca	tagctgtttc	ctgtgtgaaa	ttgttatccg	ctcacaattc	1140
cacacaacat	acgagccgga	agcataaagt	gtaaagcctg	gggtgcctaa	tgagtgaagt	1200
aactcacatt	aattgcgttg	cgtcactgctc	ccgctttcca	gtcgggaaac	ctgtcgtgcc	1260
agctgcatta	atgaatcggc	caacgcgcgg	ggagaggcgg	tttgogtatt	gggcgtctct	1320
ccgcttcctc	gctcactgac	tcgctgcgct	cggctggtcg	gctgcggcga	gcggtatcag	1380
ctcactcaaa	ggcggtaata	cggttatcca	cagaatcagg	ggataacgca	ggaaagaaca	1440
tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	ggccgcggtg	ctggcggttt	1500
tccatagggt	ccgccccct	gacgagcatc	acaaaaatcg	acgctcaagt	cagagggtggc	1560
gaaacccgac	aggactataa	agataccagg	cgtttccccc	tggaagctcc	ctcgtgcgct	1620
ctcctgttcc	gacctgtcc	cttaccggat	acctgtccgc	ctttctccct	tcgggaagcg	1680
tggcgttttc	tcatagctca	cgtgtaggt	atctcagttc	gggtgtaggtc	gttcgctcca	1740
agctgggctg	tgtgcacgaa	cccccggttc	agcccgaccg	ctgcgcctta	tccggtaact	1800
atcgtcttga	gtccaacccg	gtaagacacg	acttatcgcc	actggcagca	gccactggta	1860
acaggattag	cagtagggag	gtgctacaga	gtgctacaga	gttcttgaag	tggtggccta	1920
actacggcta	cactagaagg	acagtatttg	gtatctgcgc	tctgctgaag	ccagttacot	1980
tcggaaaaaag	agttggtagc	tcttgatccg	gcaaaacaaac	caccgctggg	agcgggtgggt	2040
ttttgttttg	caagcgcag	attacgcgca	gaaaaaaaagg	atctcaagaa	gatccctttga	2100
tcttttctac	gggggtctgac	gctcagtgga	acgaaaaactc	acgttaagggg	attttgggtca	2160
tgagattatc	aaaaaggatc	ttcacctaga	tccttttaaa	ttaaaaatga	agtttttaaa	2220
caatcataag	tatatattgag	taaacttggt	ctacagatta	ccaatgctta	atcagtgagg	2280
cacctatctc	agcgatctgt	ctatttcggt	catccatagt	tgccctgactc	cccgtcgtgt	2340

```

agataactac gatacgggag ggcttaccat ctggcccccag tgctgcaatg ataccgcgag 2400
acccacgctc accgggtcca gatttatcag caataaacca gccagccgga agggccgagc 2460
gcagaagtgg tcttgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag 2520
ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt tggtgccatt gctacaggca 2580
tcgtgggtgc acgctcgctc tttggatggt cttcattcag ctccggttcc caacgatcaa 2640
ggcgagttac atgatcccc atgttgtgca aaaaagcggg tagctccttc ggtcctccga 2700
tcgttgtcag aagtaagttg gccgcagtgt tatcactcat gggtatggca gcactgcata 2760
attctcttac tgctatgcca tccgtaagat gcttttctgt gactggtgag tactcaacca 2820
agtcattctg agaatagttg atgcggcgac cgagtgtctc ttgcccggcg tcaatacggg 2880
ataataccgc gccacatagc agaactttaa aagtgtcat cattggaaaa cgttcttcgg 2940
ggcgaaaaact ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg 3000
cacccaactg atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaaacag 3060
gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatac 3120
tcttcctttt tcaatattat tgaagcattt atcaggggta ttgtctcatg agcggataca 3180
tatttgaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt ccccgaaaag 3240
tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa aataggcgta 3300
tcacgaggcc ctttcgtctc gtcggtttcg gtgatgacgg tgaaaacctc tgacacatgc 3360
agctcccgga gacggtcaca gcttgtctgt aagcggatgc cgggagcaga caagcccgtc 3420
agggcgcgtc agcgggtgtt ggcgggtgtc ggggctggct taactatgcg gcacagatgc 3480
agattgtact gagagtgcac catatgcggt gtgaaatacc gcacagatgc gtaaggagaa 3540
aataccgcat caggcgccat tcgccattca ggcgtgcgca ctggtgggaa gggcgatcgg 3600
tcggggcctc ttcgctatta cgccagctgg cgaagggggg atgtgtgca aggcgattaa 3660
gttggttaac gccagggttt tccagctcac gacgttgtaa aacgacggcc agtgaattcg 3720
agctgtggaa tgtgtgtcag ttaggggtgt gaaagtcccc aggcctccca gcaggcagaa 3780
gtatgcaaaag catgcatctc aattagtcag caaccagggt tggaaaagtcc ccaggctccc 3840
cagcaggcag aagtatgcaa agcatgcate tcaattagtc agcaaccata gtcccgcctc 3900
taactccgcc catcccgcct ctaactccgc ccagttccgc ccattctccg ccccatggct 3960
gactaatttt ttttatttat ggcctcctc agggcgctcc ggcctctgag ctattccaga 4020
agtagtgagg aggcgttttt ggaggctcgg taccoccttg cgctaattgt ctgttacagg 4080
tactaatac catctaagta gttgattcat agtgactgca tatgttgtgt tttacagtat 4140
tatgtagtct gttttttatg caaaatctaa ttaatatat tgatatttat atcattttac 4200
gtttctcgtt cttctttttt atactaagtt ggcattataa cttatcaatt 4260
gtttgcaacg aacagggtcac tatcagtcac aataaaatca ttatttgatt tcaattttgt 4320
cccactccct gcctctgggg ggcgcg 4346

```

<210> 114  
 <211> 3166  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> p18attBZeo Plasmid

```

<400> 114
cagttgcccg cggggtcgcg cagggcgaa ccccccccc acggctgctc gccgatctcg 60
gtcatggccg gccgggaggg gtcccggaa ttcgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggccgcg caccacacc caggccaggg tggtgtccgg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtgcgtc 240
tccacgaagt cccgggagaa cccgagccgg tcggtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagtttagt taaaaaagca ggcttcaate ctgcagagaa gcttgcatgc ctgaggtcgt 420
actctagagg atccccgggt accgagctcg aattcgtaat catggtcata gctgtttcct 480
gtgtgaaatt gttatccgct cacaattcca cacaacatac gagccggaag cataaagtgt 540
aaagcctggg gtgcctaatt agtgagctaa ctacattaa ttgctgtgag ctactgccc 600
gctttccagt cgggaaacct gtccgtgccg ctgcattaat gaatcgccca acgcgcgggg 660
agaggcggtt tgcgtattgg gcgctcttcc gcttccctcg tccactgact cctgctcgtc 720
gtcgttcggc tgccggcagc ggtatcagct cactcaaagg cggtaatatc gttatccaca 780
gaatcagggg ataacgcagg aaagaacatg tgagcaaaa gcccagcaaa ggccagggaac 840
cgtaaaaagg ccgcgttgct gggcgttttc atatggctcc gccocctga cgagcatcac 900
aaaaatcgac gctcaagtca gaggtggcga caccgcacag gactataaag ataccaggcg 960
tttccccctg gaagctccct cgtgcgctct cctgttccga ccctgcccgt taccggatag 1020
ctgtccgcct ttctcccttc ggggaagcgt gcgctttctc atagctcacg ctgtaggat 1080
ctcagttcgg tcgtccaaag tgggtgtgt tgcacgaacc ccccgttcag ccccgttcag 1140
cccgaccgct gcgccttata cggtaactat cgtcttgagt ccaaccgggt aagacacgag 1200
ttatcgccac tggcagcagc cactggtaac aggtatagca gagcgaggta tgtaggcgt 1260
gtacagagat tcttgaagtg gtggcctaac taccgctaca ctagaaggac agtatctgg 1320
atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc 1380

```

```

aaacaaacca ccgctggtag cgggtggtttt tttgtttgca agcagcagat tacgcgcaga 1440
aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc tcagtggaaac 1500
gaaaactcac gttaagggat tttgggtcatg agattatcaa aaaggatctt cacctagatc 1560
cttttaaat aaaaatgaag ttttaaatca atctaaagta tatatgagta aacttggtct 1620
gacagttacc aatgcttaat cagtgaaggca cctatctcag cgatctgtct atttcgttca 1680
tccatagttg cctgactccc cgtcgtgtag ataactacga tacgggaggg cttaccatct 1740
ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga tttatcagca 1800
ataaacagc cagccggaag ggccgagcgc agaagtggtc ctgcaacttt atccgcctcc 1860
atccagctca ttaattgttg ccgggaagct agagttaagta gttcgccagt taatagtttg 1920
cgcaacggtt ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt tggatggct 1980
tcattcagct ccggttccca acgatcaagg ctagttacat gatcccccac gttgtgcaaa 2040
aaagcggtta gtccttcggg tcctccgatc gttgtcagaa gtaagtggc cgcatgtta 2100
tactcatgg ttatggcagc actgcataat tctcttactg tcatgccatc cgtaagatgc 2160
ttttctgtga ctgggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg 2220
agttgctctt gcccgcgctc aatccgggat aataccgccc caccatagcag aactttaaaa 2280
gtgctcatca ttggaaaaacg ttcttcgggg ccaaaactct cttcagcatc accgctgttg 2340
agatccagtt cgatgttaacc cactcgtgca agggcaaaatg ccgcaaaaaa gggaataagg 2400
accagcgttt ctgggtgagc aaaaacagga actcatactc ttctttttc aatattattg tttactttc 2460
gcgacacgga aatgttgaat actcactata tttgaatgta ttagaaaaaa taaacaaata 2520
cagggttatt gtctcatgag ccgaaaagtg ccacctgacg attcaggctg cgcaactgtt 2580
ggggttccgc gctttattac caagcgaagc gccattcgcc attcaggctg cgcaactgtt 2640
gggaaggcgc atcgtgccc gctcttcgc tattacgcca gctggcgaaa gggggatgtg 2700
ctgcaaggcg attaagttgg gtaacgccag ggttttccca gtcacgacgt tgtaaaacga 2760
cggccagtc gtaatacgac tcaactaagg ccttgactag agggctcgac gtatacagac 2820
atgataagat acattgatga gtttggacaa acccaacta gaatgcagtg aaaaaaatgc 2880
tttatttgtg aaatttgtga tgctattgct ttatttgtaa ccattataag ctgcaataaa 2940
caagttgggg tgggcgaaga actccagcat gagatccccg cgtggagga ccatccagcc 3000
ggcgtcccgg aaaaacgatt ccttcatag aaggcgccgg tggaaatcgaa 3060
atctcgtagc acgtgtcagt cctgctcctc ggccacgaag tgcacg 3120

```

<210> 115  
 <211> 7600  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> p18attBZeo3'6XHS4eGFP Plasmid

```

<400> 115
cagttgcccg cggggtcgcg cagggcgaa cccccccc acggctgctc gccgatctcg 60
gtcatggccc gcccggaggc gtcccgaag ttcgtggaca cgacctccga ccactcggcg 120
tacagctcgt ccaggcccgcg caccacacac caggccaggg tggtgtccgg caccacctgg 180
tcctggaccg cgctgatgaa cagggtcacg tcgctccgga ccacaccggc gaagtcgtcc 240
tccacgaagt cccgggagaa cccgagccgg tcggtccaga actcgaccgc tccggcgacg 300
tcgcgccgcg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc cccatataatg gagttccgcg ttacataact 420
gtaatcaatt acggggtcac tagttcatag cccatataatg gagttccgcg ttacataact 480
tacggtaaat ggccgcctg gctgacccgc caacgacccc gcccattga cgtcaataat 540
gacgtatgtt cccatagtaa tggcagtaga gactttccat tgaagtgtat catatgccaa gtacgcccc 600
tttacggtaa actgccact tggcagtaga tggcagtaga gactttccat tgaagtgtat catatgccaa gtacgcccc 660
tattgacgtc aatgacggtg aatggccgct ctggcattat gcccagtaga tgaccttatg 720
ggactttcct acttggcagt acatctacgt attagtcac tctctctccc cctccccccc ctccccacc 780
tgagccccac gttctgcttc ttttgaatta ttttgtgcag cgtggggggg gggggggggg gggggcgcg 840
atatttttat ttttgaatta ttttgtgcag cgtggggggg gggggggggg gggggcgcg 900
ccaggcgggg cggggcgggg cgagggggcg gggggggggg gggggggggg gggggcgcg 960
gccaatcaga gcggcgcgct cggaaagttt cgttttatgg cggggcgggg gggggcgcg 1020
ccctataaaa agcgaagcgc gggggcgggg gggggcgggg gggggcgggg gggggcgcg 1080
ccgctccgcg ccgctccgcg ccgctccgcg ccgctccgcg ccgctccgcg ccgctccgcg 1140
gtgagcgggc gggacggccc gggacggccc gggacggccc gggacggccc gggacggccc 1200
ctcgtttctt tctgtggtt ggtcgtgggg ggtcgtgggg ggtcgtgggg ggtcgtgggg 1260
ggggggggag cggcggtggt cggcggtggt cggcggtggt cggcggtggt cggcggtggt 1320
ggcggtggt cggcggtggt cggcggtggt cggcggtggt cggcggtggt cggcggtggt 1380
gtgtgcccga ggggagcggg ggggagcggg ggggagcggg ggggagcggg ggggagcggg 1440
gaacaaaggc tgcgtgcccc ggcgtgcccc ggcgtgcccc ggcgtgcccc ggcgtgcccc 1500
ggtcgggctg gctccgtgct gggcggtgct gggcggtgct gggcggtgct gggcggtgct 1560
cgggtgcccc gctccgtgct gggcggtgct gggcggtgct gggcggtgct gggcggtgct 1620

```

ggcaggtggg	ggtgccgggc	ggggcggggc	cgcttcgggc	gggggagggc	tcgggggagg	1680
ggcgcgccgg	ccccggagcg	ccggcgccgt	tcgagggcgc	gcgagccgca	gccattgcct	1740
tttatggtaa	tcgtgcgaga	ggggcgccgg	acttcctttg	tcccaaatct	ggcggagccg	1800
aaatctggga	ggcgcccgcc	cacccctctt	agcgggcgcg	ggcgaagcgg	tgccggcccg	1860
gcaggaagga	aatggggcgg	gagggccttc	gtgcgtcgcc	gcccgcgcgt	ccccttctcc	1920
atctccagcc	tcggggcgtc	cgagggggga	cggctgcctt	cggggggggac	ggggcagggc	1980
gggggttcgg	ttctggcggt	tgaccggcgg	ctctagagcc	tctgctaacc	atgttcctgc	2040
cttcttcttt	ttcctacagc	tcctgggcaa	cgtgctgggt	gttgtgctgt	ctcatcattt	2100
tggcaaagaa	ttcgccacca	tggtagcaaa	gggcgaggag	ctgttcaccg	gggtgggtgcc	2160
catcctgggt	gagctggacg	gcgacgtaaa	cggccacaag	ttcagcgtgt	ccggcgaggg	2220
cgagggcgat	gccacctacg	gcaagctgac	cctgaagttc	atctgcacca	ccggcaagct	2280
gcccgtgccc	tggcccaccc	tcgtgaccac	cctgacctac	ggcgtgcagt	gcttcagccg	2340
ctaccccgag	cacatgaagc	agcacgactt	cttcaagtcc	gccatgcccg	aaggctacgt	2400
ccaggagcgc	accatcttct	tcaaggacga	cggcaactac	aagacccgcg	ccgaggtgaa	2460
gttcgagggc	gacaccctgg	tgaaccgcat	cgagctgaag	ggcatcgact	tcaaggagga	2520
cggcaacatc	ctggggcaca	agctggagta	caactacaac	agccacaacg	tctatatcat	2580
ggcgcacaag	cagaagaacg	gcactcaagt	gcactcaag	atccgccaca	acatcgagga	2640
cggcagcgtg	cagctcgccg	accactacca	gcagaacacc	cccatcgccg	acggccccgt	2700
gctgctgccc	gacaaccact	acctgagcac	ccagtcggcc	ctgagcaaa	accccaacga	2760
gaagcgcgat	cacatggtcc	tgctggagtt	tcaggtgcag	gcccgggatca	ctctcggcat	2820
ggacgagctg	tacaagtaag	aattcactcc	ataccactga	gctgcctatc	agaaggtggt	2880
ggctgggtgt	gccaatgccc	tggtgcacaa	gcactctgact	gatctttttc	cctctgcaa	2940
aaattatggg	gacatcatga	agccocttga	gttggaattt	tctggctaata	aaaggaaatt	3000
tattttcatt	gcaatagtgt	gttggaattt	gtatttggtt	tcactcggaa	ggacatatgg	3060
gagggcaaat	cattttaaac	atcagaatga	tataaagagg	tagagtgttg	caacatatgc	3120
catatctggg	ctggccatga	caaaggtggc	aagccttgac	tcactcagtat	atgaaacagc	3180
cccctgctgt	ccattcctta	ttccatagaa	catccctaaa	ttgaggttag	atttttttta	3240
tattttgttt	tgtgtttttt	ttttctttta	tcccagtcac	attttcttta	catgtttttac	3300
tagccagatt	tttctctctc	tcctgactac	catgcctgca	agctgtccct	cttctcttat	3360
gaagatccct	cgacctgcag	cccaagcttg	tcaaagagca	ggtcgactct	agtggatccc	3420
ccgccccgct	tccccccagt	gtctgcaggc	gcacgctgcc	gcgagaagcg	ttcagaggaa	3480
agcgatcccg	cccgtgcccg	ggagcggagc	tcgctgctgc	gcacgctgcc	ggctcgggga	3540
tgccggggga	gcccgggacc	cctggggggt	gctgtctgag	cccgacgctg	cccctagcgg	3600
gggagggagc	taattacatc	gtgtctgcag	ggggtgtcc	gctgtccccg	tgagcggatc	3660
cgccggcccc	tatccccacg	tcctcgctgc	gccccggggc	cagcgagaag	cggtcgggga	3720
aaagcgatcc	cgtgccacct	cccccgctgc	ttgggggggg	cccgacgctg	ccggctcggg	3780
gatcgggggg	gagcgccgga	ccggagcgga	ggccttgggg	gctcgctgct	gccccctagc	3840
gggggagggg	cgtaattaca	tcctgggggg	agggctcaa	gggctgtccc	cgtagagcga	3900
tccgcccggc	cgtatcccc	aggtgtctgc	cccggtgtgt	agcagcgaga	agcgttcaga	3960
ggaaagcgat	cccgtgccac	cttccccgtg	gagccccggg	ccccgcacgc	tgccggctcg	4020
gggatgcggg	gggagcgccg	gaccggagcg	ggccttgggg	cggctcgctg	ctgcccccta	4080
gcgggggagg	gacgtaatta	catccctggg	gcaggtcaa	gggggctgtc	cccgtgagcg	4140
gatccgcggc	cccgtatccc	ccaggtgtct	tgccccgggt	agagcagcga	gaagcgttca	4200
gaggaaagcg	atccccgtgc	accttccccg	cggagccccg	gtccccgcac	gctccggggt	4260
cggggatgcg	gggggagcgc	cggaccggag	ggggcttggg	ggcggtcgcg	tgctgcccc	4320
tagcggggga	gggacgtaat	tacatccctg	ctgcaggctc	gggggggctg	tccccgtgag	4380
cggatccgcg	gccccgtatc	ccccaggtgt	cgtgccccgg	aaagagcagc	gagaagcgtt	4440
cagaggaaag	cgatccccgt	ccacottccc	agcggagccc	ctgtccccgc	acgctgcccg	4500
ctcgggggat	cggggggagc	gcccggaccg	ttggggcttt	cgggcggctc	gctgctgccc	4560
cctagcgggg	gagggacgta	attacatccc	gtctgcaggc	gggggggggg	tgcccccggt	4620
agcggatccg	cggccccgta	tccccccagt	cccgtgcccc	tcaaagagca	gcgagaagcg	4680
ttcagaggaa	agcgatcccc	tgccaccttc	ggagcggagc	ggctgtcccc	gcacgctgcc	4740
ggctcgggga	tgccggggga	gcccgggacc	cctggggggg	cccgggcggc	tcgctgctgc	4800
ccccagcggg	gggagggagc	taattacatc	taatttgctg	ttgggggggg	gctgtccccg	4860
tgagcgggat	cgccggggct	caggaattcg	taatcatggt	catagctgtt	tcctgtgtga	4920
aattgtttat	cgtccacaat	tccacacac	atacagagcg	gaagcataaa	gtgtaaagcc	4980
tgggggtgct	aatgagtgag	ctaactcaca	ttaattgctg	tgccgtcact	gcccgttttc	5040
cagtgcggga	acctgtcgtg	ccagctgcac	taatgaatcg	gccaacgcgc	ggggagaggg	5100
ggttttgcgt	ttgggcgctc	ttccgcttcc	tcgctcactg	actcgctgcg	ctcggctggt	5160
cggctgcggc	gagcgggtac	agctcactca	aaaggccagc	tacgggtatc	cacagaatca	5220
ggggataacg	caggaaagaa	catgtgagca	ctcggccccc	aaaaggccag	gaaccgtaaa	5280
agggccgcgt	tgctggcggt	tttccatagg	gcgaaacccg	ctgacgagca	tcacaaaaat	5340
cgacgtccaa	gtcagaggtg	ctctcctggt	cctgagactat	aaagatacca	ggcggtttccc	5400
cctggaagct	ccctcgctgc	cgtggcgctt	ctctcatagct	cgtttaccgg	atacctgtcc	5460
gcctttctcc	cttcgggaag	caagctgggt	tgtgtgcacg	cacgctgtag	gtatctcagt	5520
tcggtgtagg	tcgttcgctc	ctatcgtctt	gagtcacaac	aaccccccg	tcagcccagc	5580
cgtgcgcct	tatccggtaa			cggtaagaca	cgacttatcg	5640

ccactggcag	cagccactgg	taacaggatt	agcagagcga	ggtatgtagg	cggtgctaca	5700
gagttcttga	agtgggtggc	taactacggc	tacactagaa	ggacagtatt	tggatatctgc	5760
gctctgctga	agccagttac	cttcggaaaa	agagttggta	gctcttgatc	cggaacacaa	5820
accaccgctg	gtagcgggtg	tttttttggt	tgcaagcagc	agattacgcg	cagaaaaaaa	5880
ggatctcaag	aagatccttt	gatcttttct	acggggtctg	acgctcagtg	gaacgaaaaa	5940
tcacgttaag	ggatttttgg	catgagatta	tcaaaaagga	tcttcacctg	gatcctttta	6000
aattaaaaat	gaagttttta	atcaatctaa	agtatatatg	agtaaaactt	gtctgacagt	6060
taccaatgct	taatcagtga	ggcacctatc	tcagcgatct	gtctatttcg	ttcatccata	6120
gttgccctgac	tccccgtcgt	gtagataact	acgatacggg	agggccttacc	atctggcccc	6180
agtgcctgcaa	tgataccgcg	agacccacgc	tcacccggctc	cagattttatc	agcaataaac	6240
cagccagccg	gaaggggccg	gcgacagaag	gggtcctgcaa	ctttatccgc	ctccatccag	6300
tctattaatt	gttgccggga	agctagagta	agtagttcgc	cagttaatag	tttgcgcaac	6360
gttggttgcca	ttgctacagg	catcgtgggt	tcacgctcgt	cgtttggtat	ggcttcattc	6420
agctccgggt	cccaacgatc	aaggcgagtt	acatgatccc	ccatgttgtg	caaaaaagcg	6480
gttagctcct	tcggtcctcc	gatcgttgtc	agaagtaagt	tggccgcagt	gttatcactc	6540
atgggttatgg	cagcactgca	taattctctt	actgtcatgc	catccgtaag	atgcttttct	6600
gtgactgggtg	agtactcaac	caagtcattc	tgagaatagt	gtatgcggcg	accgagttgc	6660
tcttgccggg	cgtcaatacg	ggataatacc	gcgccacata	gcagaacttt	aaaagtgtctc	6720
atcattggaa	aacgttcttc	ggggcgaaaa	ctctcaagga	tcttaccgct	gttgagatcc	6780
agttcgatgt	aacccactcg	tgaccccaac	tgatcttcag	catcttttac	tttcaccagc	6840
gtttctgggt	gagcaaaaaa	aggaaggcaa	aatgccgcaa	aaaagggaat	aagggcgaca	6900
cggaatggtt	gaatactcat	actcttcctt	tttcaatatt	attgaagcat	ttatcagggt	6960
tattgtctca	tgagcggata	catatttgaa	tgtatttaga	aaaataaaca	aatagggtgt	7020
ccgcgcacat	ttccccgaaa	agtcccacct	gacgtagtta	acaaaaaaa	gcccgcgcaa	7080
gcgggcttta	ttaccaagcg	aagcgccatt	cgccattcag	gctgcgcaac	tgttgggaag	7140
ggcgatcggt	tcgggcccct	tcgctattac	gccagctggc	gaaaggggga	tgtgctgcaa	7200
ggcgattaag	ttgggtaacg	ccagggtttt	cccagtcacg	acgttgtaaa	acgacggcca	7260
gtccgtaata	cgactcactt	aaggccttga	ctagagggtc	gacggtatac	agacatgata	7320
agatacattg	atgagtttgg	acaaaccaca	actagaatgc	agtgaaaaaa	atgcttttatt	7380
tgtgaaattt	tgatgcttat	tgctttattt	gtaaccatta	taagctgcaa	taaacaagtt	7440
gggggtggcg	aagaactcca	gcattgagatc	cccgcgctgg	aggatcatcc	agccggcgctc	7500
ccggaacacg	attccgaagc	ccaacctttc	atagaaggcg	gcggtggaat	cgaaatctcg	7560
tagcacgtgt	cagtcctgct	cctcggccac	gaagtgcacg			7600

<210> 116

<211> 7631

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeo5' 6XHS4eGFP Plasmid

<400> 116

cagttgcccg	ccgggtcgcg	cagggcgaa	tcccccccc	acggctgctc	gccgatctcg	60
gtcatggccg	gccccgaggg	gtcccgaag	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggcccg	caccacacc	caggccaggg	tggtgtccgg	caccacctgg	180
tccctggaccg	cgctgatgaa	cagggtcacg	tcgtcccgga	ccacaccggc	gaagtctgtc	240
tcacgaaggt	cccgggagaa	cccagaccgg	tcggtccaga	actcgaccgc	tccggcgacg	300
tcgcgcgcgg	tgagcaccgg	aacggcactg	gtcaaacttg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgatata	gaattcctgc	420
agcccccgcg	atccgctcac	ggggacagcc	cccccccaaa	gccccaggg	atgtaattac	480
gtcccccccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	gtccggcgct	540
cccccccgcat	ccccgagccg	gcagcgtgcg	gggacagccc	gggcacgggg	aaggtggcac	600
gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	cctgggggat	660
acggggccgc	ggatccgctc	acggggacag	ccccccccca	aagccccag	ggatgtaatt	720
acgtccctcc	cccgctaggg	ggcagcagcg	agccgcccgg	ggctccgctc	cggtccggcg	780
ctccccccgc	atcccccgag	cggcagcgtg	cggggacagc	ccgggacagg	ggaaggtggc	840
acgggatcgc	tttctctga	acgttctctg	ctgctctttg	agcctgcaga	cacctggggg	900
atacgggggoc	gcggatccgc	tcacggggac	agcccccccc	caaagccccc	agggatgtaa	960
ttacgtccct	ccccccttag	ggggcagcag	cgagccgccc	ggggctccgc	tccggtccgc	1020
cgctcccccc	gccccccga	gcccgcagcg	tcgggggaca	gcccggggac	gggggaaggtg	1080
gcacggggatc	gctttcctct	gaacgcttct	cgctgctctt	tgagcctgca	gacacctggg	1140
ggatacgggg	cccgcgatcc	gctcacgggg	acagcccccc	cccaaagccc	ccagggatgt	1200
aattacgtcc	ctcccccgct	agggggcagc	agcgagccgc	ccgggggctcc	gctccgggtc	1260
ggcgctcccc	ccgcaccccc	gagccggcag	cgtgcggggg	cagcccgggc	acggggaagg	1320
tggcacgggga	tcgcttttct	ctgaacgctt	ctcgctgctc	tttgagcctg	cagacacctg	1380
ggggatacgg	ggccgcggat	ccgctcacgg	ggacagcccc	cccccaaagc	ccccagggat	1440



gtaattacgt	ccctccccc	ctaggggggca	gcagcgagcc	gcccgggggct	ccgctccgggt	1500
ccggcgctcc	ccccgcaccc	ccgagccggc	agcgtgcggg	gacagcccgg	gcacgggggaa	1560
gggtggcacgg	gatcgcttcc	ctctgaacgc	ttctcgctgc	tccttgagcc	tgcagacacc	1620
tgggggatac	ggggcgcgcg	atccgctcac	ggggacagcc	ccccccaaa	gccccaggg	1680
atgtaattac	gtccctcccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	1740
gtccggcgct	ccccccgcac	ccccgagccg	gcagcgctgc	gggacagccc	gggcacgggg	1800
aagggtggcac	gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcggg	ggatccacta	gttattaata	gtaatcaatt	acgggggtcat	1920
tagttcatag	cccatatatg	gagttccgcg	ttacataact	tacggtaaat	ggcccgccctg	1980
gctgaccgccc	caacgacccc	cgcccatgga	cgtaataaat	gacgtatgtt	cccatagtaa	2040
cgcaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgcccact	2100
tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagtaaa	tgaccttatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tgggtcgagg	tgagccccac	gttctgcttc	2280
actctcccca	tctccccccc	ctccccaccc	ccaattttgt	atztatttat	tttttaatta	2340
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cgggggcgggg	2400
cgagggggcg	ggcgggggcg	ggcggaaggg	tgggtcgagg	gccaatcaga	gcggcgcgct	2460
ccgaaagt	cccttttatgg	cgagggcgcg	gcttgccctt	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtcgctg	cggtgccttc	tgacgcgctt	ccgctccgcg	ccgctccgcg	2580
ccgcccgcgc	cggtctgac	agcgcttggt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctcctccg	ggctgtaatt	tcggggaggg	ttaatgacgg	ctcgcttctt	ttctgtgggt	2700
gcgtgaaagc	cttaaagggc	gtggggagcg	ccctttgtgc	ggggggggagc	ggctcggggg	2760
gtgctgctg	gtgtgtgtgc	ggggcgcgcg	cccgctgcgg	cccgcgctgc	ccggcgcgct	2820
tgagcgctgc	gtgccccgcg	gtgccccggg	tgctgcgagg	gtgtgcgcga	ggggagcgcg	2880
gcccggggcg	gggggggtga	gcaggggggtg	tgggcgcggc	gaacaaaggc	tgctgccccg	2940
gtgtgtcgct	ctccccgagt	tgctgagcac	ggcccgctt	ggtcgggctg	taaccccccc	3000
ctgcaccccc	cggggctcgc	cggtgcgggc	ggggggtggc	ggcaggtggg	ggtgccccgc	3060
ggcgctggcg	cgcctcgggc	cggggagggc	tcgggggagg	ggcgcgcgcg	gggccccg	3120
ggggcggggc	tcgagggcg	gcgagccgca	gccattgctt	tttatggtaa	ccccggagcg	3180
ccggcgctgc	acttcctttg	tcacaaatct	ggcgagagcg	aaatctggga	ggcgccgccc	3240
ggggcgaggg	agcggggcg	ggcgaagcgg	tgccgctcgc	gcaggaagga	ggcgccgccc	3300
cacccctct	gtgcgtcgcc	gcgcgcgctt	cccttctcc	atctccagcc	aatggggcg	3360
gaggcccttc	cggctgcctt	cggggggggac	ggggcgaggg	gggggttcggc	tcggggctgc	3420
cgcaggggga	ctctagagcc	tcctgtaacc	atgttcatgc	cttcttctt	ttctgcccga	3480
tgaccggcg	cgctgctggt	gttggtgctt	gttcatctt	tgccaaagaa	ttcgccacca	3540
tcctgggcaa	gggcgagggg	ttcagcggtg	gggtgggtgc	catcctgggt	gagctgggag	3600
tggtgagcaa	cggccacaag	atctgcacca	ccggcgaggg	cgagggcgat	gccaactacg	3660
gcgacgtaaa	cctgacgttc	ggcggtgcag	gcttcagccg	gcccgtgccc	tgccccaccc	3720
gcaagctgac	cctgacctac	gccatgcccc	aaggctacgt	ctacccccgac	cacatgaagc	3780
tcgtgaccac	cttcaagtc	aagacccgcg	ccaggttgaa	ccaggagcgc	accatcttct	3840
agcacgactt	cggaactac	ggcatcgact	tcaggagga	gttcgagggc	gacacccctg	3900
tcaaggacga	cgagctgaag	agccacaacg	actatcgagg	cggaacatc	gttcgagggc	3960
tgaaccgcat	caactacaac	atccgccaca	acgccccgct	ggccgacaag	ctggggcaca	4020
agctggagta	gcagaaacac	cccatcgggc	accccaacga	cggcagcggt	cagaagaacg	4080
gcatcgaagt	ccagtcggcc	ctctcgccat	agaaggtggt	gctgctgccc	gagctcgccg	4140
accactacca	tgctggagtt	gctgcctatc	cctctgccaa	gaagcgcgat	gacaaccact	4200
acctgagcac	aattcactcc	gatctttttc	aaaggaaatt	ggacgagctg	cacatgggtc	4260
tgctggagtt	tggttcacaa	tctggctaatt	ttttttcatt	ggctgggtgtg	tacaagtaag	4320
aattcactcc	agcccttga	tcactcgga	gagggcaaat	aaattatggg	gcaatagtg	4380
tggttcacaa	ggttgaattt	tagagtttgg	catatgctgg	tattttcatt	gcatatagtg	4440
agcccttga	atcagaatga	tcataagagg	ccctgctgt	tagccagatt	tttccctcct	4500
ggttgaattt	caaagtgagg	ttccatagaa	ccctgctgt	gaagatccct	cgacctgcag	4560
atcagaatga	ttccatagaa	tttctttaa	tattttgttt	cggttaccca	gctcgaattc	4620
tcctgactac	tcctccttaa	tcctccttaa	tagccagatt	ccgctcacaa	ttccacacaa	4680
cccaagcttg	cattccctgca	tcctccttaa	gaagatccct	taatgagtg	gctaactcac	4740
gtaatcatgg	tcataagctg	agctgtccct	agaggatccc	aacctgtcgt	gccaagtgca	4800
atacagagcc	ggaaagcaca	ggctgcactc	aaattgttat	attgggcgct	cttccgcttc	4860
catacagagcc	ttgcgctcac	ttcctgtgtg	ctgggtgccc	cgagcggtat	cagctcactc	4920
ataatgtgcg	ggccaaacgc	agtgtaaaag	ccagtcggga	gcaggaagaa	acatgtgagc	4980
taaatgaatc	gactcgctgc	tgccccgctt	cgggtttgct	ttgctggcgt	ttttccatag	5040
ctcgctcact	gactcgctgc	cggggagagg	tcgggtgcgg	agtcagaggt	ggcgaaaccc	5100
aaaggcggt	atacgggttat	ccacagaatc	aggggataac	tcctggaagc	gctctcctgt	5160
aaaaggccag	caaaaggcca	ggaacggtaa	aaaggccgcg	ccctggaagc		5220
gctccgcccc	cctgacgagc	atcacaaaaa	tcgagaggt	tcctcgctgc		5280
gacaggacta	taaagatacc	aggcggtttcc	ccctggaagc			5340
						5400
						5460



tccgaccctg	ccgcttaccg	gatacctgtc	cgcctttctc	ccttcgggaa	gcgtggcgct	5520
ttctcatagc	tcacgctgta	ggatatctcag	ttcgggtgtag	gtcgttcgct	ccaagctggg	5580
ctgtgtgcac	gaaccccccg	ttcagccccga	ccgctgcgcc	ttatccggta	actatcgtct	5640
tgagtccaac	ccggtaagac	acgacttatc	gccactggca	gcagccactg	gtaacaggat	5700
tagcagagcg	aggatatgtag	gcgggtgctac	agagttcttg	aagtgggtggc	ctaactacgg	5760
ctacactaga	aggacagtat	ttggatatctg	cgctctgctg	aagccagtta	ccttcggaaa	5820
aagagttggg	agctcttgat	ccggcaaaaa	aaccaccgct	ggtagcggtg	gtttttttgt	5880
ttgcaagcag	cagattacgc	gcagaaaaaa	aggatctcaa	gaagatcctt	tcatcttttc	5940
tacgggggtc	gacgctcagt	ggaacgaaaa	ctcacgttaa	gggatttttg	tcatgagatt	6000
atcaaaaagg	atcttcacct	agatcctttt	aaattaaaaa	tgaagtttta	aatcaatcta	6060
aagtataat	gagtaaactt	ggctcgacag	ttaccaatgc	ttaatcagtg	aggcacctat	6120
ctcagcgatc	tgtctatttc	gttcatccat	agttgcctga	ctccccgtcg	tgtagataac	6180
tacgatacgg	gaggggcttac	catctggccc	cagtgctgca	atgataccgc	gagacccacg	6240
ctcaccggct	ccagattttat	cagcaataaa	ccagccagcc	ggaaggggccc	agcgagaaag	6300
tggtcctgca	acttttatccg	cctccatcca	gtctattaat	tgttgccggg	aagctagagt	6360
aagtagttcg	ccagtttaata	gtttgcgcaa	cgttgttgcc	attgctacag	gcctcgtggg	6420
gtcacgctcg	tcgtttggtg	tggtcttcatt	cagctccggt	tcccaacgat	caaggcgagt	6480
tacatgatcc	cccattgttg	gcaaaaaaagc	ggtagctcc	ttcggtcctc	cgatcgttgt	6540
cagaagtaag	ttggccgcag	tgttatcact	catggttatg	gcagcactgc	ataattctct	6600
tactgtcatg	ccatccgtaa	gatgcttttc	tgtgactggg	gagtactcaa	ccaagtccatt	6660
ctgagaatag	tgtatgcggc	gaccgagttg	ctcttgcccg	gcgtcaatac	gggataatac	6720
cgcgccacat	agcagaactt	taaaagtgtc	catcattgga	aaacgttctt	cggggcgaaa	6780
actctcaagg	atcttaccgc	tgttgagatc	cagttcgatg	taaccactc	gtgcacccaa	6840
ctgatcttca	gcctctttta	ctttaccag	cgtttctggg	tgagcaaaaa	caggaaggca	6900
aaatgccgca	aaaaagggaa	taagggcgac	acggaaaatgt	tgaatactca	tactcttcct	6960
ttttcaatat	tattgaagca	tttatcaggg	ttattgtctc	atgagcggat	acatatattga	7020
atgtatttag	aaaaataaac	aaataggggt	tccgcgaca	tttccccgaa	aagtgccacc	7080
tgacgtagtt	aacaaaaaaa	agcccgcgca	agcgggcttt	attaccaagc	gaagcgccat	7140
tcgccattca	ggctgcgcaa	ctgttgggaa	gggcgatcgg	tgcgggcctc	ttcgctatta	7200
cgccagctgg	cgaaaagggg	atgtgctgca	aggcgattaa	gttgggtaac	gccagggttt	7260
tcccagtcac	gacgttgtaa	aacgacggcc	agtccgtaat	acgactcact	taaggccttg	7320
actagagggt	cgacggtata	cagacatgat	aagatacatt	gatgagtttg	gacaaaccac	7380
aactagaatg	cagtgaaaaa	aatgctttat	ttgtgaaatt	tgtgatgcta	ttgctttatt	7440
tgtaaccatt	ataagctgca	ataaacaagt	tgggggtggc	gaagaactcc	agcatgagat	7500
ccccgcgctg	gaggatcatc	cagccggcgt	cccggaatac	gattccgaag	cccaaccttt	7560
catagaaggc	ggcgggtgaa	tcgaaatctc	gtagcacgtg	tcagtcctgc	tcctcggcca	7620
cgaagtgcac	g					7631

<210> 117  
 <211> 4615  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> p18attBZeo6XHS4 Plasmid

<400> 117						
cagttgcccg	ccgggtcgcg	cagggcgaa	tccgcgcgc	acggctgctc	gccgatctcg	60
gtcatggccg	gcccggaggc	gtcccgaag	ttcgtggaca	cgacctccga	ccactcggcg	120
tacagctcgt	ccaggccgcg	caccacacac	caggccaggg	tgttgctcgg	caccacctgg	180
tccctggaccg	cgctgatgaa	cagggtcacg	tcgctccgga	ccacacccgc	gaagtcgtcc	240
tcacgaagt	ccggggagaa	cccgagccgg	tcgggtccaga	actcgaccgc	tccggcgacg	300
tgcgcgcgcg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgcatgc	ctgcaggtcg	420
actctagtgg	atcccccgcc	ccgtatcccc	caggtgtctg	caggctcaaa	gagcagcgag	480
aagcgttcag	aggaaaagcga	tcccgtgcc	ccttccccgt	gcccgggctg	tccccgcacg	540
ctgccggctc	ggggatgcgg	ggggagcgcc	ggacggagc	ggagccccgg	gcggctcgct	600
gctgccccct	agcgggggag	ggacgtaatt	acatccctgg	gggctttggg	ggggggctgt	660
ccccgtgagc	ggatccgcgg	ccccgtatcc	cccagggtgc	tgcaggctca	aagagcagcg	720
agaagcgttc	agagggaagc	gatccccgtc	cacctcccc	gtgcccgggc	tgtccccgca	780
cgctgcccgc	tcggggatgc	ggggggagcg	ccggaccgga	gcggagcccc	gggcggctcg	840
ctgctgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	ggggggggct	900
gtccccgtga	gcggtaccgc	ggccccgtat	ccccagggtg	tctgcaggct	caaagagcag	960
cgagaagcgt	tcagaggaaa	gcgatccccg	gccaccttcc	ccgtgcccg	gctgtccccg	1020
cacgctgcgc	gctcggggat	gcggggggag	agccggagcc	gagcggagcc	ccggggcggt	1080
cgctgctgcc	ccctagcggg	ggagggagct	atctacatcc	ctgggggctt	tggggggggg	1140
ctgtccccgt	gagcggatcc	gcggccccgt	atcccccagg	tgtctgcagg	ctcaaagagc	1200

```

agcgagaagc gttcagagga aagcgatccc gtgccacctt ccccggtgcc gggctgtccc 1260
cgacacgtgc cggctcgggg atgccccggg agcgccggag cggagcggag ccccgggcgg 1320
ctcgctgctg ccccttagcg ggggagggag gtaattacat ccctggggggc tttggggggg 1380
ggctgtcccc gtgagcggat ccgcggcccc gtatccccca ggtgtctgca ggctcaaaga 1440
gcagcgagaa gcgttcagag gaaagcgatc ccgtgccacc ttccccgtgc ccgggctgtc 1500
ccccacgctc gccggctcgg ggatgcgggg ggagcggcgg accggagcgg agccccgggc 1560
ggctcgtcgc tgccccctag cgggggaggg acgtaattac atccctgggg gctttggggg 1620
ggggctgtcc ccgtgagcgg atccgcggcc ccgtatcccc caggtgtctg caggctcaaa 1680
gagcagcgag aagcgttcag aggaaagcga tccccgtcca ccttccccgt gcccggtgtg 1740
tccccgcacg ctgccggctc ggggatgcgg ggggagcggc ggaccggagc ggagccccgg 1800
gcggctcgtc gctgccccct agcgggggag ggacgttaatt acatccctgg gggctttggg 1860
ggggggctgt ccccgtagac ggatccgcgg ggctgcagga attcgtaatc atcgtaatc 1920
ctgtttcctg tgtgaaattg ttatccgctc tgcctaataa acaacatacg agccggaagc 1980
ataaagtgtg aagcctgggg ttatccgctc tgcctaataa tcacattaat tgcgttgccg 2040
tactgccccg ctttccagtc gggaaacctg tcgtgccagc tgcattaatg aatcgcccaa 2100
cgcgcgggga gaggcggttt gcgtattggg cgctcttccg cttcctcgct cactgactcg 2160
ctgctgtcgg tegtccggct gcggcgagcg gtatcagctc actcaaaggc ggtaatcgg 2220
ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag 2280
gccaggaacc gtaaaaaggc cgcgttgctg cggtttttcc ataggctccg cccccctgac 2340
gagcatcaca aaaatcgacg ctcaagttag aggtggcgaa acccgacagg actataaaga 2400
taccaggcgt ttccccctgg ctcaagttag gtgctccgac ctgctccgct 2460
accggatacc tgteccgctt tctcccttcg ggaagcgtgg cgctttctca tagctcacgc 2520
tgtaggtatc ccgcttcagg cgccttatcc ggtaccagc tgggctgtgt gcacgaaccc 2580
ccgcttcagg tatcgccact ggcagcagcc ggtaactatc gtcttgagtc caaccgggta 2640
agacacgact ctacagagtt ctgaagtgg tggcctaact actggtaaca agcgaggtat 2700
gtagggcgtg tctgcgtct gctgaagcca tttacacctc acggctacac tagaaggaca 2760
gtatttggtg aacaaaccac cgtggtagc ggtggttttt ttgtttgcaa tggtagctct 2820
tgatccggca aaaaaggatc tcttctatct tttctacggg gattatcaaa atatgagtaa 2880
acgcgcagaa aaaaactcac ttaagggatt ttttaaatca tctaaagtat gatctgtcta 2940
cagtggaaacg aaaaactcac ttaagggatt ttttaaatca tctaaagtat gatctgtcta 3000
acctagatcc ttttaaatca ttaagggatt ttttaaatca tctaaagtat gatctgtcta 3060
acttggtctg acagttacca ccatagttgc tgcaatgata agtgaggcac taactacgat acgggagggc 3120
tttcggtcat ccatagttgc tgcaatgata agtgaggcac taactacgat acgggagggc 3180
ttaccatctg gccccagtgc agccggaagg gccgagcgca cagctcacc ggctccagat 3240
ttatcagcaa taaaccagcc taattgttgc cgggaagcta gagtaagtag tgcacttta 3300
tccgcctcca tccagtctat tgccattgct cgggttccaa ttgcagagc ttccgccagt 3360
aatagtttgc cactcatggt cgccttcggt tatggcagca ctgcataatt gagttacatg atccccatg 3420
ggatgtgctt cattcagctc ctccttcggt cgtatcaagg cagttacatg taagttggcc 3480
ttgtgcaaaa aagcggttag ctccttcggt cgtatcaagg cagttacatg taagttggcc 3540
gcagtgttat cactcatggt ctccttcggt cgtatcaagg cagttacatg taagttggcc 3600
gtaagatgct tttctgtgac tatggcagca ctgcataatt cagttacatg taagttggcc 3660
cggcgaccga gttgctcttg cccggcgta ctccttcggt cagttacatg taagttggcc 3720
actttaaaag tgctcatcat tggaaaacgt tcttcggggc tcttcggggc ccaactgatc 3780
ccgctgttga gatccagttc gatgtaaccc actcgtgcac ggcaaaatgc cgcatactct 3840
tttactttca ccagcgttcc tgggtgagca aaaacaggaa ctcatactct ttgaatgtat ttagaaaaat 3900
ggaataaggg cgacacggaa atgttgaata ctcatactct ttgaatgtat ttagaaaaat 3960
agcattttat aggggtattg totcatgagc ggatacatat cacttaagtc agttaacaaa 4020
aaacaaatag ggggtccgcg cccggcgta cccggcgta cccggcgta cccggcgta 4080
aaaaagcccc cgaagcggg cccggcgta cccggcgta cccggcgta cccggcgta 4140
gcaactgttg ggaagggcga ttaagttggg taacgccagg cctcttcgct attacgccag 4200
ggggatgtgc tgcaaggcga taatcgact cacttaagtc gttttccag gttttccag 4260
gtaaaacgac ggccagtcg taatcgact cacttaagtc gttttccag gttttccag 4320
tatacagaca tgataagata cattgatgag tttggacaaa ccaacttag aatgcagtga 4380
aaaaaatgct ttatttgtga aatttgtgat gctattgctt ttttgtaac cattataagc 4440
tgcaataaac aagttggggg gggcgagaa ctccagcatg agatccccgc gctggaggat 4500
catccagccg gcgtcccgga aaacgattcc gaagcccaac ctttcataga agggcgcggt 4560
ggaatcgaaa tctcgtagca cgtgtcagtc ctgctcctcg gccacgaagt gcacg 4615

```

<210> 118  
 <211> 17384  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pFK161 Plasmid

<400> 118

gcgacacgagg	gagcttccag	ggggaaaacgc	ctgggtatcct	tatagtcctg	tcgggggtttc	60
gccacctctg	acttgagcgt	cgatttttgt	gatgctcgtc	agggggggcgg	agcctatgga	120
aaaacgccag	caacgcggcc	tttttacggg	tcctggccct	ttgctggcct	tttgctcaca	180
tggtctttcc	tgcggttatcc	cctgattctg	tggaataaccg	tattaccgcc	tttgagttag	240
ctgataccgc	tcgccgcagc	cgaacgaccg	agcgacgcga	gtcagtgagc	gaggaagcgg	300
aagagcgctg	acttccgcgt	ttccagactt	tacgaaacac	ggaaaccgaa	gaccattcat	360
gttggttgc	aggtcgcaga	cgttttgcag	cagcagtcgc	ttcacgttcg	ctcgcgatc	420
ggtagattcat	tctgctaacc	agtaaggcaa	ccccgccagc	ctagccgggt	cctcaacgac	480
aggagcacga	tcatgcgcac	ccgtcagatc	cagacatgat	aagatacatt	gatgagtttg	540
gacaaaccac	aactagaatg	cagtgaaaaa	aatgctttat	ttgtgaaatt	tgtgatgcta	600
ttgctttatt	tgtaaccatt	ataagctgca	ataaacaagt	taacaacaac	aattgcattc	660
attttatgtt	tcaggttcag	ggggaggtgt	gggaggtttt	ttaaagcaag	taaaacctct	720
acaaatgtgg	tatgcttgat	tatgatctct	agtcaaggca	ctatacatca	aatattcctt	780
attaacccct	ttacaaatta	aaaagctaaa	gtacacaaat	ttttgagcat	agttattaat	840
agcagacact	ctatgcctgt	gtggagtaag	aaaaaacagt	atgttatgat	tataactggt	900
atgcctactt	ataaagggtta	cagaatattt	ttccataatt	ttcttgtata	gcagtgccgc	960
tttttccttt	gtggtgtaaa	tagcaaaagca	agcaagagtt	ctattactaa	acacagcatg	1020
actcaaaaaa	cttagcaatt	ctgaaggaaa	gtccttgggg	tcttctacct	ttctcttctt	1080
ttttggagga	gtagaatggt	gagagtcagc	agtagcctca	tcactactag	atggcatttc	1140
ttctgagcaa	aacagggtttt	cctcattaaa	ggcattccac	cactgctccc	attcatcagt	1200
tcctataggt	ggaatctaaa	atacacaac	aattagaatc	agtagtttaa	cacattatac	1260
acttaaaaat	tttatatttta	ccttagagct	ttaaatctct	gtaggtagtt	tgtccaatta	1320
tgctcacacca	cagaagtaag	gttccctcac	aaagatccgg	accaaagcgg	ccatcgtagc	1380
tccccactcc	tgtagttcgg	gggcatggat	gcgcgagtag	ccgctgctgg	tttccctggat	1440
gccgacggat	ttgcaactgcc	ggtagaactc	gcgaggtcgt	ccagcctcag	gcagcagctg	1500
aaccaactcg	cgagggggatc	gagcccgggg	tgggcgaaaga	actccagcat	gagatccccg	1560
cgctggagga	tcactccagcc	ggcgctccgg	aaaacgattc	cgaagcccaa	cctttcatag	1620
aaggcggcgg	tggaatcgaa	atctcgtgat	ggcaggttgg	gcgtcgcttg	gtcgggtcatt	1680
tcgaacccca	gagtcctcgct	cagaagaact	cgtaagaag	gcgatagaag	gcgatgcgtt	1740
gcgaatcggg	agcggcgata	ccgtaaaagca	cgaggaagcg	gtcagcccat	tcgccgccaa	1800
gctcttcagc	aatatcacgg	gtagccaacg	ctatgtcctg	atagcgggtc	gccacacca	1860
gccggccaca	gtcgatgaat	ccagaaaagc	ggccattttc	caccatgata	ttcggcagcc	1920
aggcatcgcc	atgggtcacg	acgagatcct	ccagctcggg	atgcgcgcct	tgagcctggc	1980
gaacagttcg	gctggcgcg	gcccctgatg	ctcttcgtcc	agatcatcct	gatcgacaag	2040
accggcttcc	atccgagtag	gtgctcgctc	gatgcgatgt	ttcgcttggt	ggtcgaatgg	2100
gcagtgagga	ggatccaagc	tatgcagcgc	ccgcatgtga	tcagccatga	tggaactttt	2160
ctcgccagga	gcaaggtgag	atgacaggag	atcctgcccc	ggcacttcgc	ccaatagcag	2220
ccagtccttt	cccgttccag	tgacaacgct	gagcacagct	gcgcaaggaa	cgcccgctcg	2280
ggccagccac	atgcttcgct	ctgcctcgct	ctgcagttca	ttcagggcac	cgacaggttc	2340
ggctcttgaca	aaaagaaccg	ggcgccccctg	cgctgacagc	cggaacacgg	cggcatacga	2400
gcagccgatt	gtctgtttgt	cccagtcata	gcccgaatagc	ctctccaccc	aagcggccgg	2460
agaacctcgc	tggaatccat	cttggttcaat	catgcgaaac	gatcctcatc	ctgtctcttg	2520
atcagatctt	gatccccctgc	gccatcagat	ccttgccggc	aagaaagcca	tccagtttac	2580
tttgacgggc	ttcccaacct	taccagaggg	ggccccagct	ggcaattccg	gttcgcttgc	2640
tgctccataaa	accgccacgt	ctagctatcg	ccatgtaagc	ccactgcaag	ctacctgctt	2700
tctcttttgcg	ccttgcgttt	cccttgtcca	gatagcccag	tagctgacat	tcactccggg	2760
tcagcacctg	ttctcgggac	tggttttcta	cgtgttccgc	ttccttttagc	agcccttgcg	2820
ccctgagtg	ttcgccgagc	tttgcaaaaag	cctaggcctc	cctaggcctc	caaaaaagcc	2880
tcctcactac	ttctggaata	gctcagaggc	cgaggcggcc	taaataaaaa	aaatttagtca	2940
gccatggggc	ggagaatggg	cggaactggg	cgaggttagg	ggcgggatgg	gcccaggttag	3000
gggcgggact	atggttgctg	actaattgag	atgcattgct	tgcatacttc	tgctgctggg	3060
ggagcctggg	gactttccac	acctgggtgc	tgactaattg	agatgcatgc	tttgcatact	3120
tctgcctgct	ggggagcctg	gggaactttcc	acaccctaac	tgacacacat	tcacagccg	3180
gatctgcagg	acccaacgct	gcccagagatg	cgccgcgtgc	ggctgctgga	gatggcgagc	3240
gcgatggata	tggttctgcca	aggggttggtt	tgccgattca	cagttctccg	caagaattga	3300
ttggctccaa	ttcttgagat	ggtgaatccg	ttagcgaggt	ggcgccggct	tccattcagg	3360
tcgaggtggc	ccggctccat	gcaccgcgac	gcaacgcggg	gaggcagaca	aggtataggg	3420
cggcgccctac	aatccatgcc	aaccgcgttc	atgtgctcgc	cgaggcgcac	aaatcgccgt	3480
gacgatcagc	ggtccaatga	tcgaagttag	gctggtaaga	gcccgcgagcg	atccttgaag	3540
ctgtccctga	tggtcgtcat	ctacctgcct	ggacagcatg	gctgcaaacg	cgccatcccg	3600
atgcgcgcgg	aagcgagaag	aatcataatg	gggaaggcca	tcagcctcgc	cgctcgcaac	3660
gccagcaaga	cgtagcccag	cgcgtcgggc	cgccatgcgc	gcgataatgg	cctgcttctc	3720
gccgaaacgt	ttggtggcgg	gaccagtgc	gaaggcttga	gcgagggcgt	gcaagattcc	3780
gaataccgca	agcgacaggg	cgatcatcgt	cgcgctccag	cgaaagcggg	cctcgccgaa	3840
aatgacccag	agcgtgccc	gcaacctgtcc	tacgagttgc	atgataaaga	agacagtcac	3900
aagtgcgggc	acgatagtc	tgccccgcgc	ccaccggaag	gagctgactg	gggtgaaggc	3960
tctcaagggc	atcggtcgac	gctctccctt	atgcgactcc	tgcattagga	agcagcccg	4020

tagtaggttg	aggcgggtga	gcaccgcccgc	cgcaagggaat	ggtgcatgca	aggagatggc	4080
gccccacagt	cccccgccca	cgggcctgcc	accataccca	cgccgaaaca	agcgctcatg	4140
agccccaagt	ggcgagcccg	atottcccca	tccgtgatgt	cgccgatata	ggcgccagca	4200
accgcacctg	tggcgccggg	gatgccggcc	acgatgcgtc	cggcgtagag	gatcttggca	4260
gtcacagcat	gcgcataatc	atgcttcgac	catgcgctca	caaagtaggt	gaatgcgcaa	4320
tgtagtaccc	acatcgtcat	cgctttccac	tgtctcgcgc	aataaagatg	gaaaatcaat	4380
ctcatggtaa	tagtccatga	aaatccttgt	attcataaat	cctccaggta	gctatatgca	4440
aattgaaaca	aaagagatgg	tgatctttct	aagagatgat	ggaatctccc	ttcagtatcc	4500
cgatgggtcaa	tgcgctggat	atgggataga	tgggaatatg	ctgattttta	tgggacagag	4560
ttgcgaactg	ttcccaacta	aaatcatttt	gcacgatcag	cgcactacga	actttaccca	4620
caaatagtca	ggtaatgaat	cctgatataa	agacaggttg	ataaatcagt	cttctacgcg	4680
catcgcacgc	gcacaccgta	gaaagtcttt	cagttgtgag	cctgggcaaa	ccgttaactt	4740
tccgcgcttt	tgtctgtgca	caggctcacc	tctaaaagga	aataaatcat	gggtcataaa	4800
attatcacgt	tgtccggcgc	ggcgacggat	gttctgtatg	cgctgttttt	ccgtggcgcg	4860
ttgctgtctg	gtgatctgcc	ttctaaatct	ggcacagccg	aattgcgcga	gcttgggttt	4920
gctgaaacca	gacacacagc	aactgaatac	cagaaagaaa	atcactttac	ctttctgaca	4980
tcagaagggc	agaaatttgc	cgttgaacac	ctggtcaata	cgctgttttg	tgagcgagca	5040
tattgcgctt	cgatgacgct	ttggcgttgag	attgatacct	ctgctgcaca	aaaggcaatc	5100
gacgagctgg	accagcgcat	tccgtgacacc	gtctccttcg	aacttattcg	caatggagtg	5160
tcattcatca	aggacgcccc	tatcgcaaat	ggtgctatcc	acgcagcggc	aatcgaaaca	5220
cctcagccgg	tgaccaatat	ctacaacatc	agccttggtg	tccagcgtga	tgagccagcg	5280
cagaacaagg	taaccgtcag	tgccgataag	ttcaaagtta	aacctgggtg	tgataccaac	5340
attgaaacgt	tgatcgaaaa	cgcgctgaaa	aacgctgctg	aatgtgcggc	gctggatgtc	5400
acaaagcaaa	tggcagcaga	ccccgggtggt	atggatgaac	tggcttccta	tgtccgcacg	5460
gccatcatga	tggaatgttt	tttgcgggtc	gttatctggc	agcagtgccg	tcgatagtat	5520
gcaattgata	attattatca	gggttttcgc	ctttccggcg	atccgccttg	ttacggggcg	5580
gcgacctcgc	gggttttcgc	tatttatgaa	atattttccg	tttaaggcgt	ttccgtttct	5640
cttcgtcata	acttaatgtt	tttattttaa	ataccctctg	aaaagaaagg	aaacgcagag	5700
tgctgaaagc	gagctttttg	gcctctgtgc	tttcccttct	ctgtttttgt	ccgtggaatg	5760
aacaatggaa	gtcaacaaaa	agcagctggc	tgacattttc	ggtgcgagta	tccgtaccat	5820
tcagaactgg	caggaacagg	gaatgcccgt	tctgcgaggc	ggtggcaagg	gtaatgaggt	5880
gccttatgac	tctgcgcgcg	tcataaaatg	gtatgccgaa	agggatgctg	aaattgagaa	5940
cgaacacat	ggcgaggagg	ttgaagaact	gcggcaggcc	agcgaggcag	atccacagga	6000
cgggtgtggg	cgccatgatc	gcgtagtcga	tagtggtccc	aagtagcgaa	gcgagcagga	6060
ctgggcggcg	gcaaagcggt	cggacagtgc	tccgagaacg	ggtgcgcata	gaaattgcat	6120
caacgcata	aggcccggca	gcacgccata	gtgactggcg	atgctgtcgg	aatggacgat	6180
atcccgcga	aggatgacga	gtaccggcat	aaccaagcct	atgcctacag	catccagggt	6240
gacggtgcgc	aactgtgata	tgagcgcatt	gtagatttct	atacacggtg	cctgactcgc	6300
ttagcaattt	gcccgcgctc	aactaccgca	ttaaagctta	tcgatgataa	gcggtcaaac	6360
atgagaattc	tgggtgtgtg	ttctcgttct	gccagcgggc	cctcgtctct	ccaccccatc	6420
cgtctgcgcg	tcagtgtgtg	aaggcagggg	tgccgctctc	cggcccgacg	ctgcccgcgc	6480
cgcacttttc	cacttttggtc	gcgtgtgtct	tgaccatggt	tgaggcgccc	ggttgtgtcc	6540
tcacgtgttt	tgcataccct	gtgtctcgct	tgtgtgcacg	cccagagtcg	gtggatgtgg	6600
ccggtggcgt	tggagcgttc	tcccgtctgg	tcttaggtgc	cgctgtttct	tgtaagcgtc	6660
gaggtgtccc	tgccctgggtg	caggttttgtc	gctccgtctg	ctgcttctga	gctgggtggg	6720
gcgctcccca	gcccgtgaga	tgccctccggt	ggagagaagg	gctgtgtgcc	ttcccgtttg	6780
tgtctgagaa	agggcgcccac	gggggggtcga	gtacgcctgt	aggggcaaga	cccccttctc	6840
tcgtcgggtg	ggggttgga	cccgcgacta	gagactcatt	gcgtagggct	gggtgctgagc	6900
ggtcgcggct	gcttttcggg	agtttctcga	gcagggtctc	gctttcccgt	ggggagcttt	6960
gagagccctg	gaagagaaac	gggaccggtt	cgcagacccc	ccctgtccgc	ggatgctcag	7020
aatgcccttg	ggtttccctg	tgtgctcgctc	gcacacccc	cccgcgcggt	cgcccgcgcg	7080
ttggtctctc	tgggtccgctc	ccgcccctcag	gcatgcaccc	tctctcgggtg	gcccggggctc	7140
gtcgggggtt	cgggcttctt	acgggtctcga	tgagaaagtt	tccttctcta	gctatcttcc	7200
ggaaagggtg	gaggtctctt	cgcgccgcga	gggtctctct	ccgaatgggc	ccctggaggg	7260
ctcgcgccct	gaccgcctcc	cgcgccgcga	gcggttgctc	tctcgtctac	cgcggcccg	7320
ggcctccccg	ctccgagttc	ggggagggat	cacgcggggc	agagcctgtc	tgtcgtccctg	7380
ccgttgctgc	ggagcatgtg	gctcgggttg	tgtggtgggt	ggctggggag	agggctccgt	7440
gcacaccccc	gcgtgcgcgt	actttcctcc	cctcctgagg	gcccgcgtgc	ggacgggggtg	7500
tgggtaggcg	acgggtgggct	cccgggtccc	caccgcgtct	cccgtgcctc	acccgtgctc	7560
tcogtcgcgt	gcgtccctct	cgctcgcgtc	cacgactttg	gcccgtcccg	cgacggcgcc	7620
ctgcgcgcgc	cgtgggtgcgt	gctgtgtgct	tctcgggctg	tgtggttgtg	tcgcctcgcc	7680
cccccttccc	cgcggcagcg	ttcccacggc	tggcgaaatc	gcccggagtc	tccttccccct	7740
cctcgggggtc	gagaggggtcc	gtgtctggcg	ttgatgtatc	tcgctctcgg	ggacggggacc	7800
gttctgtggg	agaaaggctg	ttggccgcgt	ccggcgcgac	gtcggacgtg	gggacccact	7860
gcccgtcggg	gggtcttcgtc	ggtaggcac	ggtgtgtcgg	catcggtctc	tctctcgtgt	7920
cggtgtcgcc	tcctcgggct	ccgggggggc	cgtcgtgttt	cgggtcggct	cggcgctgca	7980
gggtgtgggg	gactgctcag	gggagtggtg	cagtggtgatt	cccgcgggtt	ttgcctcgcg	8040

tgcctgacc	gggtccgacgc	ccgagcgggtc	tctcgggtccc	ttgtgaggac	ccccctccgg	8100
gagggggccc	tttcggccgc	ccttgccgtc	gtcgccggcc	ctcggtctgc	tggtgcgttc	8160
ccccctcccc	gtcgcgcgca	gcccgtcttt	tttctctct	ccccctctct	cctctgactg	8220
acccgtggcc	gtgctgtcgg	accccccgca	tgggggcggc	cgggcacgta	cgcgtccggg	8280
cgggtcaccgg	gggtcttgggg	ggggggccgag	gggtaaagaaa	gtcgggtcgg	cgggcgggag	8340
gagctgtggt	ttggaggggcg	tcccggcccc	gcggccgtgg	cgggtgtctg	cgcgggtcttg	8400
gagagggctg	cgtgcgaggg	gaaaaggttg	ccccgcgagg	gcaaaggga	agaggctagc	8460
agtgggtcatt	gtcccgcagg	tggtggtggtc	tggtggccga	ggtgcgtctg	ggggggtcgt	8520
ccggccctgt	cgtccgtcgg	gaaggcgcgt	gttggggcct	gcccggagtgc	cgagggtgggt	8580
accctggcgg	tgggattaac	cccgcgcgcg	tgccccgggtg	tggcgggtggg	ggctccgggtc	8640
gatgtctacc	tccctctccc	cgagggtctca	ggctctctcc	gcgcggggtc	tcggccctcc	8700
cctcgttctc	ccctctcgcg	gggttcaagt	cgctcgtcga	cctccccctcc	tccgtccttc	8760
catctctcgc	gcaattggcgc	cgcccagagt	cacgggtgggt	tcgtcctccg	cctccgcttc	8820
tcgcccgggg	ctggccgctg	tccgggtctct	cctgcgccgac	ccccgttggc	gtgggtctct	8880
ctcgccgggt	tcgcccagctc	ctggcttcgc	ccggagggtc	agggggcttc	ccgggttcccc	8940
gacgttgccg	ctcgtctgctg	tggtcttggg	ggggggccgc	tgccggcctcc	gcccggccgt	9000
gcgccccctg	cgcacccgcg	gggtgtgcgg	ttcgccgcg	gggtcagttgg	gcccgtctct	9060
tggtgcgcgt	cgggagcgtg	tccgcctcgc	ggcggtctaga	cgcgggtgtc	gcccgggtcc	9120
gacgggtggc	ctatccaggg	ctcgcccccg	ccgacccccg	cctgcccgtc	ccgggtggtgg	9180
tcgttgggtg	ggggagtgaa	tggtgctacc	ggtcattccc	tcccgcgtgg	tttgactgtc	9240
tcgcccgggt	cgcgcttctc	tttccgccaa	ccccccacgcc	aaccaccac	cctgctctcc	9300
cggccccgggtg	cgggtcgacgt	tccgggtctc	ccgatgcccga	gggggttcggg	atttgtgccg	9360
gggagcggagg	ggagagcggg	taagagaggt	gtcgagagc	tgccccgggg	cgacgctcgg	9420
gttgggtttg	ccgcgtgcgt	gtgctcgcgt	acgggttttg	tcggaccccc	acgggggtcgg	9480
tccggccgca	tgcactctcc	cgttccgcgc	gagcgcgccg	ccgggtcacc	cccgggtttgt	9540
cctcccgcga	ggctctccgc	gcgcgcgcgc	tcctctctct	ctctcgcgct	ctctgtccc	9600
cctgggtcctg	tcccaccccc	gcgctccgc	tcgcgcttcc	ttacctgggt	gatcctgcca	9660
ggtagcatat	gcttgtctca	aagattaagc	catgcatgtc	taagtacgca	cggccgggtac	9720
agtgaactg	cgaatggctc	attaaatcag	ttatggttcc	tttggtcgtc	cgctcctctc	9780
ctacttggat	aactgtggta	attctagagc	taatacatgc	cgcagggcgc	tgacccccct	9840
tcccgggggg	ggatgcgtgc	atttatcaga	tcaaaaccaa	cccgggtgagc	tccctcccgg	9900
ctccggcccg	gggtcgggcg	ccggcggtct	ggtgactcta	gataacctcg	ggccgatcgc	9960
acgcccccg	tgggcggcgc	gacccattcg	aacgtctgcc	ctatcaactt	tcgatggtag	10020
tcgcccgtgcc	taccatgggtg	accacgggtg	acggggaatc	aggggttcgat	tccggagagg	10080
gagcctgaga	aacggctacc	acatccaagg	aaggcagcag	gcgcgcaaat	taccactcc	10140
cgacccgggg	aggtagtgac	gaaaaataac	aatacaggac	tcttccgagg	ccctgtaatt	10200
ggaatgagtc	cactttaaat	cctttaacga	ggatccattg	gagggcaagt	ctgggtgccag	10260
cagcccggtg	aattccagct	ccaatagcgt	atattaaagt	tgctgcagtt	aaaaagctcg	10320
tagttggatc	ttgggagcgg	tcgggcgggtg	gcgcgcgagg	cgagtcaccc	ccggtcccc	10380
ccccttgccct	ctcggcgccc	cctcgatgct	cttagctgag	tgccccgcgg	ggcccgaagc	10440
gtttactttg	aaaaaattag	agtgttcaaa	gcaggcccga	gcccgcctga	taccgcagct	10500
aggaataatg	gaggttctat	tcgttgggt	tcgttgggt	ttcggaactg	aggccatgat	10560
taagagggac	ggccgggggg	attcgatttg	cgccgctaga	ggtgaaattc	ttggaccggc	10620
gcaagacgga	ccagagcgaa	agcatttgcc	aagaatgttt	tcattaatca	agaacgaaag	10680
tcggagggttc	gaagacgatc	agataccgtc	gtagttccga	ccataaaacga	tgccgactgg	10740
cgatgcggcg	gcgttatctc	catgaccgcg	cgggcagctt	ccgggaaacc	aaagtctttg	10800
ggttccgggg	ggagtatggt	tgcaaaagctg	aaacttaaa	gaattgacgg	aagggcacca	10860
ccaggagtgg	gcctgcggct	taatttgact	caacacggga	aacctcacc	ggcccggaga	10920
cggacaggat	tgacagattg	atagctcttt	ctcgattccg	tgggtgggtg	tgcatggccg	10980
ttcttagttg	gtggagcgat	ttgtctgggt	aattccgata	acgaacgaga	ctctggcatg	11040
ctaactagtt	acgcgacccc	cgagcgggtc	gcgtccccca	acttcttaga	gggacaagtg	11100
gcgttcagcc	acccgagatt	gagcaataac	aggtctgtga	tgcccttaga	tgtccggggc	11160
tgcacgcgcg	ctacactgac	tggtcagcg	tggtcctacc	ctgcgcggc	agggcggggt	11220
aaccggttga	accccatctg	tgatggggat	cggggattgc	aattattccc	catgaacgag	11280
gaattcccag	taagtgcggg	tcataagctt	gcgttgatta	agtcctgcc	ctttgtacac	11340
accgcccgtg	gctactaccg	attggatggt	ttagtggagg	cctcggtatc	gccccgcggg	11400
ggtcggccca	cggccctggc	ggagcgctga	gaagacggtc	gaacttgact	atctagagga	11460
agtaaaagtc	gtaacaaggt	ttccgtaggt	gaacctgcgg	aaggatcatt	aaacggggaga	11520
ctgtggagga	gcggcgggcg	ggcccgcctc	ccccgtcttg	tggtgtgctc	cgcggggagg	11580
cgcgtcgctg	ccgggtcccg	tcgcccgcg	gtggagcgag	gtgtctggag	tgaggtgaga	11640
gaaggggtgg	gtggggctcg	tctgggtccg	tctgggaccg	cctccgattt	ccccctcccc	11700
tccctctccc	ctcgtccggc	tctgacctgc	ccacctacc	gcggcgggcg	ctgctcgcgg	11760
gcgtcttccc	tctttcccg	ccggctcttc	cggtctacg	agggggcggt	cgtcgttacg	11820
gggtttttgac	ccgtccccgg	ggcgttcggg	cgtcggggcg	cgcgctttgc	tctcccgcca	11880
cccatccccg	ccgcgggtct	ggcttttcta	cgttgggtgg	ggcggttggt	gcgtgtgggg	11940
ggatgtgagt	gtcgcgtgtg	ggctcgcgg	tcccgatgcc	acgcttttct	ggcctcgcgt	12000
gtcctccccg	ctcctgtccc	gggtacctag	ctgtcgcgtt	ccggcgcgga	ggtttaaggga	12060

ccccgggggg	gtcgccctgc	cgccccagg	gtcggggggc	gggtggggccc	gtaggggaagt	12120
cggtcggttcg	ggcgggtctc	cctcagactc	catgaccctc	ctccccccgc	tgccgcgcgtt	12180
cccaggggcgg	cggtcgtgtg	gggggggtgga	tgtctggagc	cccctcgggc	gccgtggggg	12240
cccgaccgcg	gccgcccgt	tgccccatth	ccgcgggtcg	gtcctgtcgg	tgccgggtcgt	12300
gggttccccgt	gtcgtttccc	tggttttccg	ctcccgacc	tttttttttc	ctccccccca	12360
cacgtgtctc	gtttcgttcc	tgctggcccg	cctgaggcta	cccctcggtc	catctgttct	12420
cctctctctc	cggggagagg	agggcggtgg	tcgttggggg	actgtgccgt	cgtcagcacc	12480
cgtgagttcg	ctcacacccg	aaataccgat	acgactctta	gcgggtggatc	actcggctcg	12540
tgctgcgatg	aagaacgcag	ctagctgcga	gaattaatgt	gaattgcagg	acacattgat	12600
catcgacact	tcaaacgcac	ttgcggcccc	gggttcctcc	cggggctacg	cctgtctgag	12660
cgtcggtttg	cgatcaatcg	cgtaaccgcg	tgcggtgggt	gctgcgcggc	tgggagtttg	12720
ctcgaggggc	caacccccca	accgcgggtc	ggccctccgt	ctcccgaaat	tcagacgtgt	12780
gggcgggttgt	cggtgtggcg	cgcgcgcggc	cgctcgcgag	cctgggtctcc	cccgcgcac	12840
cgcgctcgcg	gcttcttccc	gctccgcggt	tcocgcctc	gcccggtgcac	cccggctctg	12900
gcctcgcgct	ggcgccctccc	ggaccgctgc	ctcaccagtc	tttctcggtc	ccgtgccccg	12960
tgggaaccca	ccgcgcccc	gtggcgcccc	gggggtgggg	cgctccgcac	tgctctggtc	13020
gaggttggcg	gttgagggtg	tgctgcgcgc	gaggttgggt	tcgggtccct	gcggcgccgg	13080
ggttgtcggg	gtggcggtcg	acgaggggcg	gtcggctcgcc	tgcggtgggt	gtctgtgtgt	13140
gtttgggtct	tgcgctgggg	gaggcggggt	cgaccgctcg	cgggggtggc	gcggtcgccc	13200
ggcgccgcgc	accctccggc	ttgtgtggag	ggagagcgag	ggcgagaacg	gagagaggtg	13260
gtatcccccg	tggcgttgcg	agggaggggt	tggcgtcccg	cgctccgtccg	tccctccctc	13320
cctcggtggg	cgccttcgcg	ccgcacgcgg	ccgctagggg	cggtcggggc	ccgtggcccc	13380
cgtggctctt	cttcgtctcc	gcttctcctt	cacccgggcg	gtaccgcgtc	cgcgcccgcc	13440
ccgcgggacg	ccgcggcgct	cgtgcgcgca	tgcgagtcac	ccccgggtgt	tgcgagttcg	13500
gggagggaga	gggcctcgct	gaccggttgc	gtcccggctt	ccctgggggg	gaccggcgct	13560
ctgtgggctg	tgcttcccgg	gggttgcgtg	tgagttaagt	cctccacccc	cgccgcggcg	13620
ccctcccgcc	ggcctctcgg	ggacccccctg	agacgggtcg	ccggctcgct	ctcccgctgc	13680
cgcgggtggc	gtctctttcc	cgcgcgcctc	ctcgctctct	tcttcccgcg	gctggggcgg	13740
tgtccccctt	ttctgaccgc	gacctcaagt	cagacgtggc	gaccgcgtga	atttaagcat	13800
attagtcagc	ggaggaaaag	aaactaacca	ggattccctc	agtaacggcg	agtgaacagg	13860
gaagagccca	gcgcgcgaatc	ccgcgcgcgc	gtcgcggcgt	gggaaatgtg	gcgtacggaa	13920
gaccaactcc	ccggcgccgc	tcgtgggggg	cccaagtcct	tctgatcgag	gcccgcggcg	13980
tggacgggtg	gaggcccggt	gcggccccgg	cgcgcggggc	tcgggtcttc	ccggagtcgg	14040
gtttgcttggg	aatgcagccc	aaagcgggtg	gtaaaactcca	tctaaggcta	aataccggca	14100
cgaagccgat	agtcaaacag	taccgtaagg	acgggttgaa	aagaactttg	aagagagagt	14160
tcaagagggg	gtgaaaccgt	taagaggtaa	aggggtgggg	tcgcgcagat	ccgcccggag	14220
gattcaaccc	ggcggcgcgc	gtccggccgt	gcccgggtgt	cccgccggat	ctttcccgcg	14280
ccccggttct	cccgaccctc	ccaccgcgcg	gtcgttcccc	tcttctccc	cgcgtccgct	14340
gcctccggcg	cggggcgcg	gggttgggtg	gggttggcg	cgcgggcggg	gcccgggggtg	14400
gggtcggcg	gggaccgccc	ccggccggcg	accggccgcc	gcccggcgca	cttccaccgt	14460
ggcggtgcgc	cgcgaccggc	tcggggacgc	ccgggaaggc	ccggtgggga	aggtggctcg	14520
ggggggggcg	cgcgtctcag	ggcgcccgca	accacctcac	cccagtggtt	acagccctcc	14580
ggccgcgcct	tcgcccgaatc	ccggggccga	ggaagccaga	taccgcctcg	cgcgtctccc	14640
ctctcccccc	gtccgcctcc	cggggcgggc	tgggggtggg	ggccggggcg	cccccccac	14700
ggcgcgaccg	ctctcccacc	cccctccgtc	gcctctctcg	gggcccgggtg	gggggggggg	14760
cggactgtcc	ccagtgcgcc	ccgggcgtcg	tcgcgcgcgt	gggtcccggg	gggaccgtcg	14820
gtcacgcgtc	tcccgcagaa	gccgagcgca	cggggtcggc	ggcgatgtcg	gctaccacc	14880
cgacccgtct	tgaaacacgg	accaaggagt	ctaaccgctg	cgcgagtcag	gggctcgtcc	14940
gaaagccgcc	gtggcgcaat	gaagggtgaag	ggccccgcgc	ggggggccga	ggtgggatcc	15000
cgaggccctct	ccagtccgcc	gagggcgcac	caccggcccg	tctcgcccgc	cgcgcggggg	15060
aggtggagca	cgagcgtagc	cgttaggacc	cgaaagatgg	tgaactatgc	ttgggcaggg	15120
cgaagccaga	ggaaactctg	gtggaggtcc	gtagcggtcc	tgacgtgcaa	atcggtcgtc	15180
cgacctgggt	atagggggca	aagactaatc	gaaccatcta	gtagctgggt	ccctccgaag	15240
tttccctcag	gatagctggc	gctctcgctc	ccgacgtacg	cagttttatc	cggtaaagcg	15300
aatgattaga	gggtcttggg	ccgaaacgat	ctcaacctat	tctcaaacct	taaatgggta	15360
agaagcccg	ctcgctggcg	tggagccggg	cgtggaatgc	gagtgcctag	tggggccactt	15420
ttggtaagca	gaactggcgc	tgcgggatga	accgaacgcc	gggttaaggc	gcccgatgcc	15480
gacgctcatc	agaccccaga	aaaggtgttg	gttgatatag	acagcaggac	ggtggccatg	15540
gaagtccgaa	tccgctaagg	agtgtgtaac	aactcacctg	ccgaatcaac	tagccctgaa	15600
aatggatggc	cgggcccgat	cgggcccgat	cccggccgtc	gcccgcagtcg	gaacgggaacg	15660
ggacggggagc	ggccgcgaat	tcttgaagac	gaaagggcct	cgtgatacgc	ctatttttat	15720
aggttaagt	catgataata	atggttttctt	agacgtcagg	tggcactttt	cggggaaatg	15780
tgccgggaac	ccctatttgt	ttatttttctt	aaatatgtat	aaatatgtat	ccgctcatga	15840
gacaataacc	ctgataaatg	cttcaataat	attgaaaaag	gaagagtatg	agtattcaac	15900
atttccgtgt	cgccttattt	cccttttttct	cggcattttg	cttccgtgtt	ttgctcacc	15960
agaaacgctg	gtgaaagtta	aagatgctga	agatcagttg	ggtgcacgag	tgggtttacat	16020
cgaactggat	ctcaacagcg	gtaagatcct	tgagagtttt	cgccccgaag	aacgttttcc	16080



```
aatgatgagc actttttaag ttctgctatg tggcgcggtg ttatcccgtg ttgacgcccgg 16140
gcaagagcaa ctcgggtcgcc gcatacacta ttctcagaat gacttgggtg agtactcacc 16200
agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca gtgctgccat 16260
aaccatgagt gataacactg cggccaactt acttctgaca acgatcggag gaccgaagga 16320
gctaaccgct tttttgcaca acatggggga tcatgtaact cgcttgatc gttgggaacc 16380
ggagctgaat gaagccatac caaacgcaga gcgtgacacc acgatgcctg cagcaatggc 16440
aacaacgttg cgcaactat taactggcga actacttact ctagcttccc ggcaacaatt 16500
aatagactgg atggaggcgg ataaagtgc aggccactt ctgctgcctg cccttccggc 16560
tggctgggtt attgctgata aatctggagc cgggtgagcgt ggggtctcgg gtatcattgc 16620
agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacga cggggagtc 16680
aatagactgg atggaggcgg atagacagat cgctgagata ggtgcctcac tgattaagca 16740
ttggtaactg tcagaccaag ttactcata tatactttag attgatttaa aacttcattt 16800
ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca aaatccctta 16860
acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag gatcttcttg 16920
agatcctttt tttctgcgcg taatctgctg cttgcaaaaa aaaaaaccac cgctaccagc 16980
ggtaggtttt ttgcccggatc aagagctacc aactctttt ccgaaggtaa ctggcttcag 17040
cgagcgcag ataccataaata ctgtccttct agttagccg tagttaggcc accacttcaa 17100
gaactctgta gcaccgccta catacctcgc tctgctaata ctgttaccag tggctgctgc 17160
cagtgccgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac cggataagga 17220
gcagcgttcg gggtgaacgg ggggttcgtg cacacagccc agcttggagc gaacgacct 17280
caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc cgaagggaga 17340
aaggcggaca ggtatccggt aagcggcagg gtcggaacag gaga 17384
```

<210> 119  
<211> 2814  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> pLITMUS38 Plasmid

```
<400> 119
gttaactacg tcaggtggca cttttcgggg aaatgtgcgc ggaaccccta tttgtttatt 60
tttctaaata cattcaata tgtatccgct catgagacaa taaccctgat aaatgcttca 120
ataatattga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc ttattccctt 180
ttttgcggca ttttgccttc ctgtttttgc tcacccagaa acgctgggtg aagtaaaaga 240
tgctgaagat cagttgggtg cagcagtggtg ttacatcgaa ctggatctca acagcggtaa 300
gatccttgag agtttttcgcc ccgaagaacg ttctccaatg atgagcactt ttaaagtctt 360
gctatgtggc ccggtattat ccggtgttga ccccgggcaa gagcaactcg gtcgcgcgat 420
acactattct cagaatgact tgggttgagta ctccaccagtc acagaaaagc atcttacgga 480
tggcatgaca gtaagagaat tatgcagtgc tgccataacc atgagtata acactcgccc 540
caacttactt ctgacaacga tcggaggacc gaaggagcta accgcttttt tgcacaacat 600
gggggatcat gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag ccataccaaa 660
cgacgagcgt gacaccacga tgccgtgtagc aatggcaaca acgttgcgca aactattaac 720
tggcgaaacta cttactctag cttcccgcca acaatttaata gactggatgg aggcggataa 780
agttgcagga ccacttctgc gctcggccct tccggctggc tggtttattg ctgataaatc 840
tggagccggg gagcgtgggt ctcgcgggtat cattgcagca ctggggccag atggttaagc 900
ctcccgatc gtagtattct acacgacggg gagttaggca actatggatg aacgaaatag 960
acagatcgct gagataggtg cctcactgat taagcattgg taactgtcag accaagttta 1020
ctcatatata ctttagattg atttaccccg gttgataatc agaaaagccc caaaaacagg 1080
aagattgtat aagcaaatat ttaaattgta aacgttaata ttttgttaaa attcgcgtta 1140
aatttttgtt aaatcagctc attttttaac caataggccg aaatcggcaa atcccttat 1200
aaatcaaaag aatagcccgga gatagggttg agtgtgttcc cagtttggaa caagagttca 1260
ctattaaaga acgtggactc caacgtcaaa cggcgaaaaa cgtctatca gggcgatggc 1320
ccactacgtg aaccatcacc caaatcaagt tttttggggt cgaggtgccg taaagcacta 1380
aatcggaacc ctaaggaggag ccccgcattt agagcttgac ggggaaagcg aacgtggcga 1440
gaaagggaagg gaagaaagcg aaaggagcgg gcgctagggc gctggcaagt gtacgggtca 1500
cgctgcgcgt aaccaccaca cccgcgcgcg ttaatgcgcc gctacagggc gcgtaaaagg 1560
atctaggtga agatcctttt tgataatctc atgacaaaaa tcccttaacg tgagttttcg 1620
ttccactgag cgtcagaccc cgtagaaaaa atcaaggat cctcttgaga tctttttttt 1680
ctgcgcgtaa tctgctgctt gcaaacaaaa aaaccaccgc taccagcggg ggtttgtttg 1740
ccggatcaag agctaccaac tctttttccg aaggtaactg gcttcagcag agcgcagata 1800
ccaaatactg ttcttctagt gtacgcgtag taggaccacc acttcaagaa ctctgtagca 1860
ccgcctacat acctcgtctc gctaactctg ttaccagtgg ctgctgccag tggcgataag 1920
tcgtgtctta ccgggttggc ctcaagacga tagttaccgg ataaggcgca gcggtcgggc 1980
tgaacggggg gttcgtgcac acagcccgac cgtgagcgaa cgacctacac cgaactgaga 2040
tacctacagc gtgagctatg agaaagcgcc acgcttcccg aagggagaaa ggccgacagg 2100
```

tatccggtaa	gcgccagggt	cggaacagga	gagcgacga	gggagcttc	agggggaaac	2160
gcctggatc	tttatagtc	tgctgggtt	cgccacctc	gacttgagc	tcgattttt	2220
tgatgctcg	cagggggcg	gagcctatg	aaaaacgca	gcaacgcgc	ctttttacg	2280
ttcctggcc	tttgctggc	ttttgctc	atgtaatgt	agttagctc	ctcattagg	2340
acccaggct	ttacacttt	tgcttcggc	tcgtatgtt	tggtgaatt	tgagcggat	2400
acaatttc	acaggaaac	gctatgaca	tgattacgc	aagctacgt	atacgaact	2460
ctagtgggc	ccgtgcaat	gaagccggc	ggcgccaag	ttctctgc	gatatctga	2520
tcacgaatt	cgctagctt	ggccgtgac	cgctccgga	tgtagaggc	tgcgctgac	2580
ctctagctc	ggccttaag	gagtcgtat	acggactgg	cgctcttta	caacgctcg	2640
actgggaaa	ccctggcgt	acccaactt	atcgccctt	agcacatcc	cctttcgca	2700
gctggcgtaa	tagcgaag	gcccgcacc	atcgccctt	ccaacagtt	cgcgccctg	2760
atggcgaat	gcgcttcgt	tggtataaa	gcccgcctc	gcgggcttt	tttt	2814

<210> 120

<211> 2847

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attB Plasmid

<400> 120

gttaactac	tcaggtggc	cttttcggg	aaatgtgcg	ggaacccta	tttgtttat	60
tttctaaat	cattcaaata	tgtatccgt	catgagaca	taaccctgat	aaatgcttc	120
ataatattg	aaaaggaaga	gtatgagat	tcaacattt	cggtgcgcc	ttattccct	180
ttttgcggc	ttttgcctt	ctgtttttg	tcacccagaa	acgtgggtg	aagtaaaag	240
tgctgaagat	cagttgggtg	cacgagtg	ttacatcgaa	ctggatctc	acagcggta	300
gatccttgag	agttttcgcc	ccgaagaac	ttctccaat	atgagcact	ttaaagtct	360
gctatgtgg	gcggtattt	cccgtgttg	gcgcgggca	gagcaactc	gtcgcgcgt	420
acactattt	cagaatgact	tggttgagta	ctcaccagtc	acagaaaag	atcttacgg	480
tggtcatgac	gtaagagaat	tatgcagtg	tgccataacc	atgagtgata	acactgcgg	540
caacttact	ctgacaacga	tcggaggacc	gaaggagcta	accgctttt	tgcaacaac	600
gggggatcat	gtaactcgcc	ttgatcgtt	ggaaccggag	ctgaatgaag	ccataccaa	660
cgacgagcgt	gacaccacga	tgctgttag	aatggcaaca	acgttgcgca	aactattaac	720
tggggaact	cttactctag	cttcccgga	acaattaata	gactggatg	aggcggata	780
agttgcagg	ccacttctgc	gctcggccct	tcgggctgg	tggtttatt	ctgataaat	840
tggaagcgg	gagcgtgggt	ctcggcggt	cattgcagca	ctggggccag	atggtaagc	900
ctcccgatc	gtagtattc	acacgacgg	gagtcaggca	actatggat	aacgaaatg	960
acagatcgct	gagataggt	ctcactgat	taagcattg	taactgtcag	accaagtta	1020
ctcatatata	ctttagattg	atttaccctg	gttgataatc	agaaaagccc	caaaaacag	1080
aagatttgtat	aagcaaatat	ttaaattgta	aacgttaata	ttttgttaaa	attcgcgtta	1140
aatttttgt	aaatcagctc	attttttaac	caataggccg	aaatcgga	aatcccttat	1200
aatcaaaaag	aatagcccga	gatagggtt	agtgtgtt	cagtttgga	caagagtcca	1260
ctattaaaga	acgtggactc	caacgtcaaa	ggcgcaaaaa	ccgtctatca	gggcgatgg	1320
ccactacgtg	aaccatcacc	caaatcaagt	tttttgggt	cgaggtgccg	taaagcacta	1380
aatcggaacc	ctaaagggag	cccccgattt	agagcttgac	ggggaaagcg	aacgtggcga	1440
gaaaggaagg	gaagaaagcg	aaaggagcgg	gcgctagggc	gctggcaagt	gtagcggta	1500
cgctgcgcgt	aaccaccaca	ccgcgcgcgc	ttaatgcgc	gctacagggc	gcgtaaaagg	1560
atctaggtga	agatcctttt	tgataatctc	atgaccaaaa	tcctttaacg	tgagttttcg	1620
ttccactgag	cgtcagaccc	cgtagaaaa	atcaaaggat	cttcttgaga	tccttttttt	1680
ctgcgcgtaa	tctgctgctt	gcaaaacaaa	aaaccaccgc	taccagcgg	ggtttgttt	1740
ccggatcaag	agctaccaac	tctttttccg	aaggtaactg	gcttcagcag	agcgcagata	1800
ccaataactg	ttcttctagt	gtagccgtag	ttaggccacc	acttcaagaa	ctctgtagca	1860
ccgcctacat	acctcgctct	gctaactctg	ttaccagtgg	ctgctgccag	tgccgataag	1920
tcgtgtctta	ccgggtttgga	ctcaagacga	tagttaccgg	ataaggcgca	gcggtcggg	1980
tgaacggggg	gttcgtgcac	acagcccagc	ttggagcgaa	cgacctacac	cgaactgaga	2040
tacctacagc	gtgagctatg	agaaaagcgc	acgcttcccg	aagggagaaa	ggcgacagg	2100
tatccggtaa	gcgccagggt	cggaacagga	gagcgacga	gggagcttc	agggggaaac	2160
gcctggatc	tttatagtc	tgctgggtt	cgccacctc	gacttgagc	tcgattttt	2220
tgatgctcg	cagggggcg	gagcctatg	aaaaacgca	gcaacgcgc	ctttttacg	2280
ttcctggcc	tttgctggc	ttttgctc	atgtaatgt	agttagctc	ctcattagg	2340
acccaggct	ttacacttt	tgcttcggc	tcgtatgtt	tggtgaatt	tgagcggat	2400
acaatttc	acaggaaac	gctatgaca	tgattacgc	aagctacgt	atacgaact	2460
ctagtgggc	ccgtgcaat	gaagccggc	ggcgccaag	ttctctgc	gattgaagc	2520
tcgcttttta	tactaaact	agcgaaatct	ggatccacga	attcgctagc	ttcgccgtg	2580
agcgctctcc	ggatgctac	gcattgcgt	accccttagt	caaggcctta	agtgatcgt	2640
attacggact	ggcgtcgtt	ttacaacgtc	gtgactggga	aaaccctggc	gttaccaca	2700



ttaatcgcc	tgcagcacat	ccccctttcg	ccagctggcg	taatagcgaa	gaggcccgca	2760
ccgatcgccc	ttcccaacag	ttgcgcagcc	tgaatggcg	atggcgcttc	gcttggtaat	2820
aaagcccgct	tcggcgggct	ttttttt				2847

<210> 121

<211> 4223

<212> DNA

<213> Artificial Sequence

<220>

<223> pLIT38attBBSRpolyA2 Plasmid

<400> 121

accatgaaaa	cattttaacat	ttctcaacaa	gatctagaat	tagtagaagt	agcgacagag	60
aagattacaa	tgctttatga	ggataataaa	catcatgtgg	gagcggcaat	tcgtacgaaa	120
acaggagaaa	tcatttcggc	agtacatatt	gaagcgtata	taggacgagt	aactgtttgt	180
gcagaagcca	ttgcgattgg	tagtgacggt	tcgaatggac	aaaaggattt	tgacacgatt	240
gtagctgtta	gacaccctta	ttctgacgaa	gtagatagaa	gtattcgagt	ggtaagtcct	300
tgtggtatgt	gtagggagtt	gatttcagac	tatgcaccag	attgttttgt	gttaatagaa	360
atgaatggca	agttagtcaa	aactacgatt	gaagaactca	ttccactcaa	atatacccca	420
aattaaaagt	tttaccatac	caagcttggc	tgctgcctga	ggctggacga	cctcgccggag	480
ttctaccggc	agtgcacatc	cgtcggcatc	caggaaacca	gcagcggcta	tcgcgcgcatc	540
catgcccccg	aactgcagga	gtggggaggc	acgatggccg	ctttggtecg	gatctttgtg	600
aaggaaacctt	acttctgtgg	tgtgacataa	ttggacaaac	tacctacaga	gatttaaagc	660
tctaaggtaa	atataaaatt	tttaagtgt	taattgtgta	aactactgat	tctaattgtt	720
tgtgtatttt	agattccaac	ctatggaact	gatgaatggg	agcagtgggtg	gaatgccttt	780
aatgaggaaa	acctgttttg	ctcagaagaa	atgccatcta	gtgatgatga	ggctactgct	840
gactctcaac	attctactcc	tccaaaaaag	aagagaaagg	tagaagaccc	caaggacttt	900
ccttcagaat	tgctaagttt	tttgagtcac	gctgtgttta	gtaatagaac	tcttgcttgc	960
tttgctattt	acaccacaaa	ggaaaaagct	gcactgctat	acaagaaaat	tatggaaaaa	1020
tattctgttaa	cctttataag	taggcataac	agttataatc	ataacatact	gttttttctt	1080
actccacaca	ggcatagagt	gtctgctatt	aataactatg	tgatagtgc	cctgactaga	1140
agcttttttaa	tttgtaaaag	ggtttaataa	gaatatttga	gcttttaaaa	acctcccaca	1200
gatcataatc	agccatacca	cattttgtaga	ggttttactt	gttggttaact	tggtttattg	1260
cctccccctg	aacctgaaac	ataaaatgaa	tgcaatttgt	gttggttaact	tggtttattg	1320
agcttataat	ggttacaaat	aaagcaatag	catcacaaaat	ttcacaaaata	aagatccaga	1380
tttcgctcaa	gttagtataa	aaaagcaggc	ttcaatcctg	cagagaagct	tgggcccgagc	1440
cggcttcaat	tgacggggcc	ccactagtga	ctcgtattac	gtagcttggc	gtaactcgag	1500
tcataagctgt	ttcctgtgtg	aaattgttat	cgctcacaaa	ttccacacaa	catacgagcc	1560
ggaagcataa	agtgtaaagc	ctgggggtgc	taatgagtga	gctaactcac	attacatgtg	1620
agcaaaaggc	cagcaaaagg	ccaggaaccg	taaaaaggcc	gcgttgctgg	cgtttttcca	1680
taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtccaga	gggtggcgaaa	1740
cccgcacagga	ctataaagat	accaggcggt	tccccctgga	agctccctcg	tgccgtctcc	1800
tgttccgacc	ctgcccgtta	ccggatacct	gtccgccttt	ctcccttcgg	gaagcgtggc	1860
gctttctcat	agctcacgct	gtaggatatc	cagttcggtg	taggtcggtc	gctccaagct	1920
gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	1980
tcttgagtc	aacccggtaa	gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	2040
gattagcaga	gcgagggtatg	taggcggtgc	tacagagttc	ttgaagtggg	ggcctaacta	2100
cggctacact	agaagaacag	tatttggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	2160
aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	gtgggttttt	2220
tgtttgcaag	cagcagatta	cgcgacagaaa	aaaaggatct	caagaagatc	ccttgatctt	2280
ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	taagggattt	tggtcatgag	2340
attatcaaaa	aggatcttca	cctagatcct	tttacgcgcc	ctgtagcggc	gcattaagcg	2400
cggcggtgtg	ggtggttacg	cgcagcgtga	ccgctacact	tgccagcgcc	ctagcgcccg	2460
ctcctttcgc	tttcttccct	tcctttctcg	ccacgttcgc	tttcccgcgc	aagctctaaa	2520
tcggggggtc	cctttagggg	tccgatttag	tgctttacgg	cacctcgacc	ccaaaaaact	2580
tgatttgggt	gatgggttcc	gtagtgggac	atcgccctga	tagacgggtt	ttcgcccttt	2640
gacgttggag	tcacagttct	ttaatagtgg	actcttggtc	caaactggaa	caacactcaa	2700
ccctatctcg	ggctattctt	ttgattttata	agggattttg	ccgatttcgg	cctatttggt	2760
aaaaaatag	ctgattttaac	aaaaatttaa	cgcgaatttt	aacaaaatat	taacgtttac	2820
aattttaaata	tttgcttata	caatcttctt	gtttttgggg	cttttctgat	tatcaaccgg	2880
ggtaaatcaa	tctaaagtat	atatgagtaa	acttggtctg	acagttacca	atgcttaact	2940
agtgaaggac	ctatctcagc	gatctgtcta	ttctggtcat	ccatagttgc	ctgaccccc	3000
gtcgtgtaga	taactacgat	acgggagggc	ttaccatctg	gccccagtg	tgcaatgata	3060
ccgcgagacc	cacgctcacc	ggctccagat	tctacagcaa	taaaccagcc	agccgggaag	3120
gccgagcgca	gaagtgggtc	tgcaacttta	ttcgcctcca	tccagttctat	taattgttgc	3180
cgggaagcta	gagttaagtag	ttcgccagtt	aatagtttgc	gcaacgttgt	tgccattgct	3240

acaggcatcg	tgggtgtcacg	ctcgtcggtt	gggtatggctt	cattcagctc	cggttcccaa	3300
cgatcaaggc	gagttacatg	atcccccatg	ttgtgcaaaa	aagcgggttag	ctccttcggt	3360
cctccgatcg	ttgtcagaag	taagttggcc	gcagtggtat	cactcatggt	tatggcagca	3420
ctgcataatt	ctcttactgt	catgccatcc	gtaagatgct	tttctgtgac	tgggtgagta	3480
tcaaccaagt	cattctgaga	atagtgatg	ggcgaccga	gttgcctctg	cccggtcgta	3540
acacgggata	ataccgcgcc	acatagcaga	actttaaaag	tgctcatcat	tggagaacgt	3600
tcttcggggc	gaaaactctc	aaggatctta	ccgctgttga	gatccagttc	gatgtaaccc	3660
actcgtgcac	ccaactgatc	ttcagcatct	tttactttca	ccagcgtttc	tgggtgagca	3720
aaaacaggaa	ggcaaaatgc	cgcaaaaaag	ggaataaggg	cgacacggaa	atgttgaata	3780
ctcatactct	tcttttttca	atattattga	agcattttatc	aggggttattg	tctcatgagc	3840
ggatacatat	ttgaatgtat	ttagaaaaat	aaacaaatag	gggttcgcgc	cacatttccc	3900
cgaaaagtgc	cacctgacgt	agttaacaaa	aaaaagcccc	ccgaagcggg	ctttattacc	3960
aagcgaagcg	ccattcgcca	ttcaggctgc	gcaactgttg	ggaagggcga	tcggtgcggg	4020
cctcttcgct	attacgcagc	ctggcgaaaag	ggggatgtgc	tgcaaggcga	ttaagttggg	4080
taacgccagg	gttttcccag	tcacgacgtt	gtaaaacgac	ggccagtcgc	taatacgact	4140
cacttaaggc	cttgactaga	gggtcgacgc	atgcctgtac	atccggagac	gcgtcacggc	4200
cgaagctagc	gaattcgtgg	atc				4223

<210> 122

<211> 2686

<212> DNA

<213> Artificial Sequence

<220>

<223> pUC18 Plasmid

<400> 122

tcgcgcggtt	cggtgatgac	gggtgaaaacc	tctgacacat	gcagctcccg	gagacgggtc	60
cagcttgtct	gtaagcggat	gccggggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcggttg	tcggggcttg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtg	180
accatgatcg	gtgtgaaata	ccgcacagat	gcgtaaggag	aaaataccgc	atcaggcgcc	240
attgccatt	caggctgcgc	aactgttggg	aaggcgatc	ggtgcggggc	tcttcgctat	300
tacgccagct	ggcgaaaagg	ggatgtgctg	caaggcgatt	aagttgggta	acgccagggt	360
tttcccagtc	acgacgttgt	aaaacgacgg	ccagtgccaa	gcttgcatgc	ctgcaggctc	420
actctagagg	atccccgggt	accgagctcg	aattcgtaat	catggtcata	gctgtttcct	480
gtgtgaaatt	gttatccgct	cacaattcca	cacaacatac	gagccggaag	cataaagtgt	540
aaagcctggg	gtgcctaatt	agtgcctaa	ctcacattaa	ttgcgttgcg	ctcactgccc	600
gctttccagt	cgggaaacct	gtcgtgccag	ctgcattaat	gaatcggcca	acgcgcgggg	660
agaggcggtt	tgcgtattgg	gcgctcttcc	gcttccctgc	tcactgactc	gctgcgctcg	720
gtcgttcggc	tgcggcgagc	ggatcagct	cactcaaagg	cggtaatacg	gttatccaca	780
gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaaag	gccagcaaaa	ggccaggaac	840
cgtaaaaagg	ccgcgttgct	ggcggttttc	cataggctcc	gccccctga	cgagcatcac	900
aaaaatcgac	gctcaagtca	gaggtggcga	aaaccgacag	gactataaag	ataccaggcg	960
tttccccctg	gaagctccct	cgtgcgctct	cctgttccga	ccctgccgct	taccggatac	1020
ctgtccgcct	ttctcccttc	gggaagcgtg	gcgctttctc	atagctcacg	ctgtagggtat	1080
ctcagttcgg	tgtaggtcgt	tcgctccaag	ctgggctgtg	tgcacgaacc	ccccgttcag	1140
cccagccgct	gcgccttatc	cggtaactat	cgtcttgagt	ccaaccgggt	aagacacgac	1200
ttatcgccac	tggcagcagc	cactggtaac	aggattagca	gagcgaggta	tgtaggcggg	1260
gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	ctagaaggac	agtatttggt	1320
atctgcgctc	tgtctgaagcc	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	1380
aaacaaaacca	ccgctggtag	cggtgggtttt	tttgtttgca	agcagcagat	tacgcgcaga	1440
aaaaaaggat	ctcaagaaga	tcttttgatc	ttttctacgg	ggtctgacgc	tcagtggaaac	1500
gaaaactcac	gttaagggat	tttggtcatg	agattatcaa	aaaggatctt	cacctagatc	1560
cttttaaatt	aaaaatgaag	ttttaaatca	atctaaaagta	tatatgagta	aacttgggtc	1620
gacagttacc	aatgcttaat	cagtgaggca	cctatctcag	cgatctgtct	atttcgttca	1680
tccatagttg	cctgactccc	cgtcgtgtag	ataactacga	tacgggaggg	cttaccatct	1740
ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	cggctccaga	tttatcagca	1800
ataaaccagc	cagccggaag	ggccgagcgc	agaagtggtc	ctgcaacttt	atccgcctcc	1860
atccagttta	ttaatgtttg	ccgggaagct	agagttaagta	gttcgccagt	taatagtttg	1920
cgcaacgttg	ttgccattgc	tacaggcatc	gtggtgtcac	gctcgtcggt	tggtagggct	1980
tcattcagct	ccgggttcca	acgatcaagg	cgagttacat	gatcccccct	gttgtgcaaa	2040
aaagcgggta	gctccttcgg	tcttcgcgat	tttgtcagaa	gtaagtgggc	cgcatgttta	2100
tactcatggg	ttatggcagc	actgcataat	tctcttactg	tcattgccatc	cgtaagatgc	2160
ttttctgtga	ctgggtgagta	ctcaaccaag	tcattctgag	aatagtgtat	gcggcgaccg	2220
agttgctctt	gcccggcgct	aatacgggat	aataccgcgc	cacatagcag	aactttaaaa	2280
gtgctcatca	ttgggaaaacg	ttcttcgggg	cgaaaaactct	caaggatctt	accgctgttg	2340
agatccagtt	cgatgtaacc	cactcgtgca	cccaactgat	cttcagcatc	ttttactttc	2400

accagcggtt	ctgggtgagc	aaaaacagga	aggcaaaatg	ccgcaaaaaa	gggaataagg	2460
gcgacacgga	aatgttgaat	actcatactc	ttcctttttc	aatattattg	aagcatttat	2520
caggggttatt	gtctcatgag	cggatacata	tttgaatgta	tttagaaaaa	taaacaaata	2580
gggggttccgc	gcacatttcc	ccgaaaagtg	ccacctgacg	tctaagaaac	cattattatc	2640
atgacattaa	cctataaaaa	taggcgtatc	acgaggccct	ttcgtc		2686

<210> 123  
 <211> 8521  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> pCXeGFPattB (6xHS4) 2 Plasmid

<400> 123						
tacggggcg	gggatccact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccataat	ggagttccgc	gttacataac	ttacggtaaa	tggcccgccct	ggctgaccgc	120
ccaacgaccc	cgcgccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtag	240
atcaagtgt	tcatatgcca	agtacgcccc	ctattgacgt	caatgacggg	aaatggcccc	300
cctggcatta	tgcccagtag	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattagtc	cgctattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctcccc	420
atctcccccc	cctccccacc	cccaattttg	tattttat	ttttttaatt	attttgtgca	480
gcgatggggg	cggggggggg	ggggggcgcg	ggcaggcggg	gcgggggcg	gagggggcg	540
ggggggggcg	aggcgagag	gtgcgggcg	agccaatcag	agcgggcgcg	tccgaaagtt	600
tccttttatg	gcgaggcggc	ggcgggcgcg	gcccataaaa	aagcgaagcg	cgcgggcggc	660
gggagtcgt	gcgttgccct	cgcccggtgc	cccgctccgc	gcccgcctcg	gcccgcctcg	720
ccggctctga	ctgaccgcgt	tactccacca	ggtgagcggg	cgggacggcc	cttctcctcc	780
gggctgta	tagcgcttgg	tttaatgacg	gctcgcttct	tttctgtggc	tgcgtgaaag	840
ccttaaagg	ctccgggagg	gccccttgtg	cgggggggag	cggtcgggg	ggtgcgtgcg	900
tgtgtgtgtg	cgtggggagc	gcccgcgtgc	gcccgcgtgc	cccgccggct	gtgagcgctg	960
cgggcgggcg	gcggggcttt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggccgggggg	1020
ggtgccccgc	gggtgcgggg	ggctgcgagg	ggaacaaagg	ctgctgtcgg	ggtgtgtgcg	1080
tgggggggtg	agcagggggg	gtggggcgcg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgtgagca	cgggccggct	tccgggtgcg	ggctccgtgc	ggggcggtgg	1200
gcggggctcg	ccgtgcgggg	cggggggtgg	cggggggtgg	gggtgcgggg	cgggggcggg	1260
ccgcctcggg	ccggggaggg	ctcgggggag	ggggcgggcg	gccccggagc	gcccggcggt	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttatggta	atcgtgagag	agggcgaggg	1380
gacttccttt	gtcccaaatc	tggcgagacc	gaaatctggg	aggcgccgcc	gcacccccct	1440
tagcgggcgc	gggcgaagcg	gtgcgggcgc	ggcaggaagg	aaatggggcg	ggaggggcct	1500
cgtgcgtcgc	cgcgcgcgcg	tccccttctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acggctgcct	ctggggggga	cgggggcagg	cggggttcgg	cttctggcgt	gtgacggcg	1620
gctctagagc	ctctgctaac	catgttcatg	ccttcttctt	tttctacag	ctcctgggca	1680
acgtgctggt	tggtgtgctg	tctcatcatt	ttggcaaaag	attcgccacc	atggtgagca	1740
agggcgagga	gctgttcacc	gggggtggtg	ccatcctggt	cgagctggac	ggcgacgtaa	1800
acggccacaa	gttcagcgtg	tccggcgagg	gcgagggcga	tgccacctac	ggcaagctga	1860
ccctgaagtt	catctgcacc	accggcaagg	tgcccgtgcc	ctggcccacc	ctcgtgacca	1920
ccctgacct	cggcgtgcag	tgcttcagcc	gctaccccga	ccacatgaag	cagcacgact	1980
tcttcaagtc	cgccatgccc	gaaggctacg	tccaggagcg	caccatcttc	ttcaaggacg	2040
acggcaacta	caagaccgcg	gcccaggtga	agttcgaggg	cgacacccct	gtgaaccgca	2100
tcgagctgaa	gggcatcgac	gtctatatca	tggccgacaa	cctggggcac	aagctggagt	2160
acaactacaa	cagccacaac	aacatcgagg	acggcagcgt	gcagaagaac	ggcatcaagg	2220
tgaacttcaa	gatccgccac	gacggccccg	tgctgctgcc	gcagctcgcc	gaccactacc	2280
agcagaacac	ccccatcggc	gaccccaacg	agaagcgcg	cgacaaccac	tacctgagca	2340
cccagtcgc	cctgagcaaa	actctcgga	tgagcagct	tcacatgggt	ctgctggagt	2400
tcgtgaccgc	cgccggggtc	cagaaggtgg	tggtgtgtgt	gtacaagtaa	gaattcactc	2460
ctcaggtgca	ggctgcctat	cctctgcca	aaaattatgg	ggccaatgcc	ctggctcaca	2520
aataccactg	agatcttttt	ttaaaggaaat	ttattttcat	ggacatcatg	aagcccttgg	2580
agcatctgac	ttctggctaa	aggacatatg	ggaggggcaa	tgcaatagtg	tggttgaatt	2640
ttttgtgtct	ctcactcgga	gcaacatgtg	cccatctgtg	tcatttaaaa	catcagaatg	2700
agtatttgg	ttagagtttg	tatgaaacag	atattttgtt	gctgccatga	acaaagggtg	2760
ctataaagag	gtcatcagta	gatttttttt	ctagccagat	tccattcctt	attccataga	2820
aaagccttga	cctgagggtta	acatgttttt	tcttctctta	ttgtgttatt	ttttcttta	2880
acatccctaa	aattttcctt	tcttctctta	tgaagatccc	tttctcctct	ctcctgacta	2940
ctccagtc	tagctgtccc	tggttctctg	gtgaaattgt	tcgacctgca	gcccagctt	3000
ggcgtaatac	tggtcatagc	taaagtgtaa	agcctggggg	tatccgctca	caattccaca	3060
caacatacga	gcccgaagca			gcctaagtga	tgagctaact	3120

cacattaatt	gcgttgccgt	cactgcccgc	tttccagtcg	ggaaacctgt	cgtgccagcg	3180
gatccgcctc	tcaattagtc	agcaaccata	gtcccgcgcc	taactccgcc	catcccgcgc	3240
ctaactccgc	ccagttccgc	ccattctccg	ccccatggct	gactaatttt	ttttatttat	3300
gcagaggccg	aggccgcctc	ggcctctgag	ctattccaga	agtagtgagg	aggctttttt	3360
ggaggctagt	ggatcccccg	ccccgtatcc	cccagggtgc	tgcaggctca	aagagcagcg	3420
agaagcgttc	agaggaaaagc	gatcccgtgc	caccttcccc	gtgcccgggc	tgtccccgcg	3480
cgctgcccgc	tcggggatgc	ggggggagcg	ccggaccgga	gaggagcccc	ggggcggtcg	3540
ctgctgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	gggggggggt	3600
gtccccgtga	gcggatccgc	ggccccgtat	cccccagggt	tctgcaggct	caaagagcag	3660
cgagaagcgt	tcagaggaaa	gcgatcccgt	gccaccttcc	ccgtgcccgc	gctgtccccg	3720
cacgtgcccg	gctcggggat	gcggggggag	ccgcccggac	gagcggagcc	ccggcggtcg	3780
cgctgctgcc	ccctagcggg	ggagggacgt	aattacatcc	ctgggggctt	tggggggggg	3840
ctgtccccgt	gagcggatcc	gcggccccgt	atcccccagg	tgtctgcagg	ctcaaagagc	3900
agcgagaagc	gttcagaggga	aagcgatccc	gtgccacctt	ccccgtgccc	gggctgtccc	3960
cgcacgcctg	cggctcgggg	atgcgggggg	agcgcgggac	cggagcggag	ccccggggcg	4020
ctcgtctgct	ccccctagcg	ggggagggag	gtaattacat	ccctgggggc	tttggggggg	4080
ggctgtcccc	gtgacgggat	ccgcggcccc	gtatccccca	gggtgtctga	ggctcaaaga	4140
gcagcgagaa	gcgttcagag	gaaagcgatc	ccgtgccacc	ttccccgtgc	ccgggctgtc	4200
ccccgcacgt	gcgggtcggg	ggatgcgggg	ggagcgcggg	accggagcgg	agccccgggg	4260
ggctcgtcgt	tgccccctag	ccggggaggg	acgtaattac	atccctgggg	gctttggggg	4320
ggggctgtcc	ccgtgagcgg	atccgcggcc	ccgtatcccc	cagggtgtct	caggctcaaa	4380
gagcagcgag	aagcgttcag	aggaaaagcga	tcccgtgcca	ccttccccgt	gccccgggct	4440
tccccgcacg	ctgcgggctc	ggggatgcgg	ggggagcgcc	ggaccggagc	ggagcccccg	4500
gcggctcgct	gctgccccct	agcgggggag	ggacgtaatt	acatccctgg	gggctttggg	4560
gggggggctg	ccccgtgagc	ggatccgcgg	ccccgtatcc	cccagggtgc	tgcaggctca	4620
aagagcagcg	agaagcgttc	agaggaaaagc	gatccccgtg	caccttcccc	gtgcccgggc	4680
tgtccccgca	cgctgcccgc	tcggggatgc	ggggggagcg	ccggaccgga	gaggagcccc	4740
gggcgggctg	ctgtgcccc	ctagcggggg	agggacgtaa	ttacatccct	gggggctttg	4800
ggggggggct	gtccccgtga	gcggatccgc	ggggctcgag	gaattcgatt	gaagcctgct	4860
tttttatact	aacttgagcg	aaatcaagct	ccaggctttt	tgcaaaaagc	taacttgttt	4920
attgcagctt	ataatgggta	caaataaagc	aatagcatca	caaatttcac	aaataaagca	4980
ttttttcac	ttgtggtttg	tccaaactca	tccaaactca	tcaatgtatc	tatatgtgc	5040
tggatccgct	gcattaatga	atcggccaac	gcgcggggag	aggcggtttg	cgtattgggc	5100
gctcttccgc	ttcctcgctc	actgactcgc	tgcgtccggt	cggtccggct	cggcgagcgg	5160
tatcagctca	ctcaaaggcg	gtaatacggg	atccacaga	atcaggggat	aaccgaggaa	5220
agaacatgtg	agcaaaaaggc	cagcaaaaag	ccaggaaccg	taaaaaaggcc	gcgttgctgg	5280
cgttttttcca	taggctccgc	ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtccga	5340
ggtggcgaaa	cccgacagga	ctataaagat	accaggcgtt	tccccctgga	agctccctcg	5400
tgcgtctcgc	tgttccgacc	ctgcgcgtta	ccggatacct	gtccgccttt	ctcccttcgg	5460
gaagcgtggc	gctttctcaa	tgctcacgct	gtaggtatct	cagttcgggt	taggtcggtc	5520
gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	5580
gtaactatcg	tcttgagttc	aacccggtaa	gcacgacctt	atcgccactg	gcagcagcca	5640
ctggtaaacg	gattagcaga	gcgaggtatg	taggcgggtg	tacagagttc	ttgaagtggg	5700
ggcctaacta	cggtacactt	agaaggacag	tatttgggat	ctgcgctctg	ctgaagccag	5760
ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	5820
gtggtttttt	tggttgcaag	cagcagatta	cgcgagaaa	aaaaggatct	caagaagatc	5880
ctttgatctt	ttctacgggg	tctgacgctc	agtggaaacg	aaactcacgt	taagggaattt	5940
tggtcagtag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	aatgaagtt	6000
ttaaataaat	ctaaagtata	tatgagtaaa	cttggtctga	cagttaccaa	tgtttaatca	6060
gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcatc	catagttgcc	tgactccccg	6120
tcgtgtagat	aactacgata	cgggagggct	taccatctgg	ccccagtgtc	gcaatgatac	6180
cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	gccggaaggg	6240
ccgagcgcag	aagtggctct	gcaactttat	ccgcctccat	ccagtctatt	aattggtgcc	6300
gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	6360
caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	ggttcccaac	6420
gatcaaggcg	agttacatga	tcccccatgt	tgtgcaaaaa	agcggttagc	tccttcgggtc	6480
ctccgatcgt	tgtcagaagt	aagtggggcg	cagtgttatc	actcatgggt	atggcagcac	6540
tgcataattc	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	6600
caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgctcaa	6660
tacgggataa	taccgcgcca	catagcagaa	ctttaaagt	gctcatcatt	ggaaaaagtt	6720
cttcggggcg	aaaactctca	aggatcttac	cgctggtgag	atccagttcg	atgtaaccca	6780
ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	6840
aaacaggaag	gcaaaaaagc	gcaaaaaagc	gaataagggc	gacacggaaa	tgttgaatac	6900
tcatactctt	cctttttcaa	tattattgaa	gcatttatca	gggttatgtg	ctcatgagcg	6960
gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	gggttccgcg	acattttccc	7020
gaaaagtgcc	acctggtcga	cggtatcgat	aagcttgata	tcgaattcct	gcagccccgc	7080
ggatccgctc	acggggagac	ccccccccca	aagccccag	ggatgtaatt	acgtccctcc	7140

```

cccgcctaggg  ggcagcagcg  agccgccccg  ggctccgcctc  cggctccggcg  ctccccccgc  7200
atccccgagc  cggcagcgtg  cggggacagc  ccgggcacgg  ggaaggtggc  acgggatcgc  7260
tttccctctga  acgcttctcg  ctgctctttg  agcctgcaga  cacctggggg  atacggggcc  7320
gcggatccgc  tcacggggac  agcccccccc  caaagcccc  agggatgtaa  ttacgtccct  7380
ccccgcctag  ggggcagcag  cgagccgccc  ggggctccgc  tccggtccgg  cgctcccccc  7440
gcatccccga  gccggcagcg  tgccggggaca  gcccgggcac  ggggaaggtg  gcacgggatc  7500
gctttctctt  gaacgcttct  cgctgctctt  tgagcctgca  gacacctggg  ggatacgggg  7560
ccgcggatcc  gctcacgggg  acagcccccc  cccaaagccc  ccagggatgt  aattacgtcc  7620
ctcccccgct  agggggcagc  agcgagccgc  ccggggtccc  gctccggtcc  ggcgctcccc  7680
ccgcaccccc  gagccggcag  cgtgccccgg  cagcccgggc  acggggaagg  tggcacggga  7740
tcgctttcct  ctgacgctt  ctcgctgctc  tttgagcctg  cagacacctg  ggggatacgg  7800
ggccgcggat  ccgctcacgg  ggacagcccc  cccccaaagc  ccccgaggat  gtaattacgt  7860
ccctcccccg  ctagggggca  gcagcgagcc  gcccggggct  gcacggggaa  tgggggatac  7920
ccccgcctcc  ccgagccggc  agcgtgcccc  gacagccggg  tgacagacac  ggggggatac  7980
gatcgctttc  ctctgaacgc  ttctcgctgc  tctttgagcc  ccccccaggg  atgtaattac  8040
ggggccggcg  atccgctcac  ggggacagcc  ccccccaaaa  gcccacaggg  gtccggcgct  8100
gtccctcccc  gctaggggg  cagcagcgag  gcagcggtgc  gggcacgggg  aaggtggcac  8160
ccccccgcat  ccccgagccg  gcagcggtgc  ggttctcgct  cctgcagaca  cctgggggat  8220
gggatcgctt  tcctctgaac  gcttctcgct  acccccccca  aagccccag  ggtatgtaatt  8280
acggggccgc  ggcagcgctc  acggggacag  agccgccccg  ggctccgctc  cgggtccggcg  8340
acgtccctcc  cccgctaggg  ggcagcagcg  cgggacagc  ccgggacagg  ggaaggtggc  8400
ctccccccgc  atccccgagc  cggcagcgtg  ctgctctttg  agcctgcaga  cacctggggg  8460
acgggatcgc  tttcctctga  acgcttctcg  8520

```

<210> 124  
 <211> 8851  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> p18EPOcDNA Plasmid

```

<400> 124
cagttgcccgg  cggggtccgc  cagggcgaa  tcccgcccc  acggctgctc  gccgatctcg  60
gtcatggccg  gcccggaggc  gtcccggaag  ttctgtgaca  cgacctccga  ccactcggcg  120
tacagctcgt  ccaggcccg  caccacacac  cagggtcacg  tggtgtccgg  caccacctgg  180
tcctggaccg  cgctgatgaa  caggggtcac  tgcgtccgga  ccacaccggc  gaagtccgtc  240
tccacgaagt  ccggggagaa  ccgagccgg  tcggtccaga  actcgaccgc  tccggcgacg  300
tcgcgcgcgg  tgagcaccgg  aacggcactg  gtcaacttgg  ccatggatcc  agatttcgct  360
caagttagta  taaaaaagca  ggcttcaatc  ctgcagagaa  gcttgataac  gaattcctgc  420
agcccccgcg  atccgctcac  ggggacagcc  cccccccaaa  gccccaggg  atgtaattac  480
gtccctcccc  cgctaggggg  cagcagcgag  ccgcccgggg  ctccgctccg  gtccggcgct  540
ccccccgcat  ccccgagccg  gcagcggtgc  gggcacgggg  gggcacgggg  aaggtggcac  600
gggatcgctt  tcctctgaac  gcttctcgct  gctctttgag  cctgcagaca  cctgggggat  660
acggggccgc  ggatccgctc  acggggacag  ccccccccca  aagccccag  ggatgtaatt  720
acgtccctcc  cccgctaggg  ggcagcagcg  agccgccccg  ggctccgctc  cgggtccggcg  780
ctccccccgc  atccccgagc  cggcagcgtg  cggggacagc  ccgggacagg  ggaaggtggc  840
acgggatcgc  tttcctctga  acgcttctcg  ctgctctttg  agcctgcaga  cacctggggg  900
atacggggcc  ttagctccct  cccccgctag  ggggcagcag  cagaccgccc  agggatgtaa  960
cgctcccccc  gcatccccga  gccggcagcg  tgccgggaca  tgagcctgca  ggggtccgc  1020
ggacgggatc  gctttcctct  gctcacgggg  gcagccgccc  cccaaagccc  ggggacggg  1080
ggatacgggg  ccgcggtacc  gctcacgggg  agcagccccc  ccggggctcc  ggggaaggtg  1140
aattacgtcc  ctcccccgct  agggggcagc  cgtgccccgg  cagcccgggc  ccagggatgt  1200
ggcgctcccc  ccgcatcccc  gagccggcag  ctcgctgctc  cccggggctc  gctccggtcc  1260
tggcacggga  tcgctttcct  ctgaacgctt  ggacagcccc  cagcccgggc  acggggaagg  1320
ggggatacgg  ggccgcggat  cccgctcacg  gcagcgagcc  tttgagcctg  cagacacctg  1380
gtaattacgt  cctccccccg  ctagggggca  agcgtgcccc  cccccaaagc  ccccgaggat  1440
ccggcgctcc  ccccgcatcc  ctctgaacgc  cccgccccgg  cccgccccgg  ccgctccggt  1500
gggtggcacg  gatcgctttc  atccgctcac  cccgccccgg  gacagccggg  gcacggggaa  1560
tgggggatac  atgtaattac  gtcctcccc  ccccgagccg  gggacagccc  tgacagacac  1620
gtccggcgct  aaggtggcac  cctgggggat  ggtatccact  gttatcaata  ccccccaggg  1680
tagttcatag  cccatatatg  gagttccgcg  ttacataact  tacggtaaat  ggcgcgctg  1740

```

gctgaccgcc	caacgacccc	cgcccattga	cgtaataat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgcccact	2100
tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagtaca	tgacctatg	ggactttcct	acttggcagt	2220
acatctacgt	attagtcac	gctattacca	tgggtcgagg	tgagccccc	gttctgcttc	2280
actctcccca	tctccccccc	ctcccccccc	ccaattttgt	atthatttat	tttttaatta	2340
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cgggcggggg	2400
cgaggggcgg	ggcggggcga	ggcggagagg	tggggcgcca	gccaatcaga	gcggcgcgct	2460
ccgaaagttt	ccttttatgg	cgaggcgggc	gcgggcgggc	ccctataaaa	agcgaagcgc	2520
gcggcgggcg	ggagtgcgtg	cgttgccctc	gccccgtgcc	ccgctccgcg	ccgcctcgcg	2580
ccgccccccc	cggtctgac	tgaccgcgtt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctcctccg	ggctgtaatt	agcgtttggt	ttaatgacgg	ctcgtttcct	ttctgtggct	2700
gcgtgaaagc	cttaaagggg	tccgggaggg	ccctttgtgc	ggggggggagc	ggctcggggg	2760
gtgcgtgcgt	gtgtgtgtgc	gtggggagcg	cccgctgcgg	cccgcgctgc	ccggcgggct	2820
tgagcgctgc	gggcgcggcg	cggggctttg	tgcgctccgc	gtgtgcgcga	ggggagcgcg	2880
gcccggggcg	gtgccccgcg	gtgccccggg	gctgcgaggg	gaacaaaggg	tgctgcgggg	2940
gtgtgtgcgt	ggggggggga	gcaggggggt	tgccggcgcg	ggtcgggctg	taaccccccc	3000
ctgcaccccc	ctccccgagt	tgctgagcac	ggcccggtct	cggttgccgg	gctccgtgcg	3060
gggcgtggcg	cggggctcgc	cgtgccccgg	gggggggtgg	ggcaggtggg	ggtgccgggg	3120
ggggcggggg	cggtctgggc	cggggagggc	tccggggagg	ggcgcgggcg	ccccggagcg	3180
ccggcggtcg	tccaggcgcg	gcgagccgca	gccattgcct	tttatggtaa	tcgtgcgaga	3240
gggcgcaggg	acttcccttg	tcccaaactc	ggcgaggccg	aaatctggga	ggcgcccgcc	3300
caccctctct	agcggcgcg	ggcgaagcgg	tgcggcgccg	gcaggaagga	aatggcgggg	3360
gagggccttc	gtgcgtcgcc	gcgcgcgcgt	ccccctctcc	atctccagcc	tcggggctgc	3420
cgcaggggga	cggtgccttt	cggggggggac	ggggcgaggc	gggggtccgg	ttctggcggt	3480
tgaccggcgg	ctctagaatg	gggggtgcac	aatgtccctg	ctggctgtgg	cttctcctgt	3540
ccctgctgtc	gctccctctg	ggcctcccag	tccctggcgc	cccaccacgc	ctcatctgtg	3600
acagccgagt	cctggagagg	tacctcttgg	aggccaagga	ggccgagaat	atcacgacgg	3660
gctgtgctga	acactgcagc	tgaatgaga	atatcactgt	ccagacacc	aaagttaatt	3720
tctatgcctg	gaagaggatg	gaggtcgggc	agcagggcgt	agaagtctgg	cagggcctgg	3780
ccctgctgtc	ggaagctgtc	ctgcggggcc	aggccctgtt	ggtcaactct	tcccagccgt	3840
gggagccctt	gcagctgcac	gtggataaag	ccgtcagtgg	ccttcgcagc	ctcaccactc	3900
tgcttcgggc	tctgggagcc	cagaaggaag	ccatctcccc	tccagatgcg	gcctcagctg	3960
ctccactccg	aacaatcact	gctgacactt	tccgcaaact	cttccgagtc	tactccaatt	4020
tcttcggggg	aaagctgaag	ctgtacacag	gggaggccct	caggacaggg	gacagtgac	4080
gtacaagtaa	gaattcactc	ctcaggtgca	gggtgcctat	cagaaggtgg	tggctgggtg	4140
ggccaatgcc	ctggctcaca	aataccactg	agatcttttt	ccctctgcca	aaaatttatg	4200
ggacatcatg	aagccccttg	agcatctgac	ttctggctaa	ttaaaggaaat	ttattttact	4260
tgcaatagtg	tgttggaaat	ttttgtgtct	ctcactcgga	aggacatatg	ggagggcaca	4320
tcatttaaaa	catcagaatg	agtatttggt	ttagagtttg	gcaacatatg	ccatatgctg	4380
gctgccatga	acaaaaggtg	ctataaagag	gtcatcagta	tatgaaacag	ccccctgctg	4440
tccattccct	attccataga	aaagccttga	cttaggttta	gatttttttt	atattttttt	4500
ttgtgttatt	ttttctttta	acatccctaa	aattttcctt	acatgtttta	ctagccagat	4560
ttttctctct	ctcctgacta	ctcccagtea	tagctgtccc	tcttctctta	tgaagatccc	4620
tccagctgca	gccccagctt	gcattgcctg	aggctgactc	tagtggatcc	cccccccggt	4680
atcccccagg	tgtctgcagg	ctcaaagagc	agcgagagag	gttcagagga	aagcgatccc	4740
gtgccacett	ccccgtgccc	gggtgttccc	cgcacgctgc	cggtcggggg	atgcgggggg	4800
agcgccggac	cggagcggag	ccccggggcg	ctcgctgctg	ccccctagcg	ggggaggggac	4860
gtaattacat	ccctgggggg	tttggggggg	ggctgtcccc	gtgagcggat	ccgcggcccc	4920
gtatccccca	gggtgtctga	gggtcaaaga	gcagcgagaa	gcgttcagag	gaaagcgatc	4980
ccgtgccacc	ttccccgtgc	ccgggctgtc	cccgcacgct	gcccggctcgg	ggatgcgggg	5040
ggagcgccgg	accggagcgg	agccccgggg	ggctcgctgc	tgccccctag	cgggggagggg	5100
acgtaattac	atccctgggg	gctttggggg	ggggctgtcc	ccgtgagcgg	atccgcggcc	5160
ccgtatcccc	cagggtgtct	caggctcaaa	gagcagcgag	aagcgttcag	aggaaaagcga	5220
tcccgctgcca	ccttccccgt	gcccgggctg	tccccgcacg	ctgcgggctc	ggggatgcgg	5280
ggggagcgcc	ggaccggagg	gggctttggg	gcccgtcgct	gctgccccct	agcggggggag	5340
ggacgtaatt	acatccctgg	gggctttggg	ggggggctgt	ccccgtgagc	ggatccgcgg	5400
ccccgtatcc	cccaggtgtc	tgcaggctca	aagagcagcg	agaagcgttc	agaggaaagc	5460
gattcccgctg	caccttcccc	gtgccccggg	tgtccccgca	cgctgccccg	tcggggatgc	5520
ggggggagcg	cgggaccgga	gcccagcccc	ggcgccctcg	ctgctgcccc	ctagcggggg	5580
agggagcgtaa	ttacatccct	gggggctttg	ggggggggct	gtccccgtga	gcggatccgc	5640
ggcccccgat	ccccacaggt	tctgcaggct	caaagagcag	cgagaagcgt	tcagaggaaa	5700
gcgatcccg	ccacacctcc	ccgtgccccg	ctgctccccg	cacgctgccc	gctcggggg	5760
gcgggggggag	cgccggaccg	gagcggagcc	ccgggcggct	cgctgctgcc	ccctagcggg	5820
ggagggacgt	aattacatcc	ctggggggct	tggggggggg	ctgtccccct	gagcggatcc	5880
gcggccccgt	atcccccagg	tgtctgcagg	ctcaaagagc	agcgagaagc	gttcagagga	5940
aagcgatccc	gtgccacctt	ccccgtgccc	gggtgtcccc	cgcacgctgc	cggtcgggg	6000

atgccccggggg	agcgccgggac	cgagagcgagg	ccccggggcg	ctcgctgctg	ccccctagcg	6060
gggggagggac	gtaattacat	ccctggggggc	tttggggggg	ggctgtcccc	gtgagcggat	6120
ccgcgggggct	gcaggaattc	gtaatcatgg	tcatagctgt	ttcctgtgtg	aaattgttat	6180
ccgctcacaa	ttccacacaa	catacgagcc	ggaagcataa	agtgtaaagc	ctgggggtgcc	6240
taatgagtga	gctaactcac	attaattgct	ttgctgtcac	tgcccgcttt	ccagtcggga	6300
aacctgtcgt	gccagctgca	ttaatgaatc	ggccaacgcg	cggggagagg	cggtttgctg	6360
attggggcgct	cttccgcttc	ctcgctcact	gactcgctgc	gctcggtcgt	tgggctgcgg	6420
cgagcggtat	cagactcact	aaaggcggtg	atacggttat	ccacagaatc	aggggataac	6480
gcaggaaaga	acatgtgagc	aaaaggccag	caaaaggcca	ggaaccgtaa	aaaggccgcg	6540
ttgctggcgt	ttttccatag	gctccgcccc	cctgacgagc	atcacaaaaa	tgcagcgtca	6600
agtcagaggt	ggcgaaaacc	gacaggacta	taaagatacc	aggcgtttcc	ccctgggaagc	6660
tccctcgtgc	gctctcctgt	tccgaccctg	ccgcttaccg	gatacctgtc	cgcctttctc	6720
ccttcgggaa	gcgtggcgct	ttctcatagc	tcacgctgta	ggtatctcag	ttcgggtgtag	6780
gtcgttcgct	ccaagctggg	ctgtgtgcac	gaaccccccg	ttcagcccga	ccgtgcggcc	6840
ttatccggta	actatcgtct	tgagccaac	ccggtaagac	acgacttatc	gccactggca	6900
gcagccactg	gtaacaggat	tagcagagcg	aggtatgtag	gcggtgctac	agagttcttg	6960
aagtgggtggc	ctaactacgg	ctacactaga	aggacagtat	ttgggtatctg	cgctctgctg	7020
aagccagttta	ccttcggaaa	aagagtttgt	agctcttgat	ccggcaaaaca	aaccaccgct	7080
ggtagcgggtg	gtttttttgt	ttgcaagcag	cagattacgc	gcagaaaaaa	aggatctcaa	7140
gaagatcctt	tgatcttttc	tacggggtct	gacgctcagt	ggaacgaaaa	ctcacgttaa	7200
gggatttttg	tcatgagatt	atcaaaaagg	atcttcacct	agatcctttt	aaattaaaaa	7260
tgaagtttta	aatcaatcta	aagtatatat	gagtaaaact	ggtctgacag	ttaccaatgc	7320
ttaatcagtg	aggcacctat	ctcagcgatc	tgtctatttc	gttcattccat	agttgcctga	7380
ctccccgtcg	tgtagataac	tacgatacgg	gagggttac	catctggccc	cagtgtgca	7440
atgataccgc	gagacccacg	ctcaccgggt	ccagatttat	cagcaataaa	ccagccagcc	7500
ggaagggccg	agcgagaaag	tggtcctgca	actttatccg	cctccatcca	gtctattaat	7560
tggtgcgggg	aagtagtagt	aagtagttcg	ccagttaata	gtttgcgcaa	cgttgttgcc	7620
attgctacag	gcactggtgt	gtcacgctcg	tcgtttggtg	tggcttcatt	cagctccggg	7680
tcccaacgat	caaggcgagt	tacatgatcc	cccatgttgt	gcaaaaaaagc	ggttagctcc	7740
ttcggtcttc	cgatcgttgt	cagaagtaag	ttggccgcag	tggttatcact	catggttatg	7800
gcagcactgc	ataattctct	tactgtcatg	ccatccgtaa	gatgcttttc	tgtgactggg	7860
gagtactcaa	ccaagtcatt	ctgagaatag	tgtatgcggc	gaccgagttg	ctcttgcccg	7920
ggtcacaata	gggcaataac	cgcgcacat	agcagaactt	taaaagtgtc	catcatttga	7980
aaacgttctt	cggggcgaaa	actctcaagg	atcttaaccg	tgttgagatc	cagttcgatg	8040
taaccacact	gtgcacccaa	ctgatcttca	gcattcttta	ctttcaccag	cgtttctggg	8100
tgagcaaaaa	caggaaggca	aaatgcccga	aaaaagggaa	taagggcgac	acggaaatgt	8160
tgaataactca	tactcttcct	ttttcaatat	tattgaagca	tttatcaggg	ttattgtctc	8220
atgagcggat	acatatattga	atgtatttag	aaaaataaaa	aaataggggt	tccgcgcaca	8280
tttccccgaa	aagtggccac	tgacgtagtt	aacaaaaaaa	agcccgccga	agcgggcttt	8340
attaccaagc	gaagcgccat	tccgcatcca	ggctgcgcaa	ctgttgggaa	gggcgatccg	8400
tgcggggcctc	ttcgctatta	cgccagctgg	cgaaaggggg	atgtgctgca	aggcgattaa	8460
ggtgggtaac	gcccagggtt	tcccagtcac	gacgttgtta	aacgacggcc	agtccgtaat	8520
acgactcact	taaggccttg	actagagggg	cgacgggtata	cagacatgat	aagatacatt	8580
gatgagtttg	gacaaaaccac	aactagaatg	cagtgaaaaa	aatgctttat	ttgtgaaatt	8640
tgtgatgcta	ttgctttatt	tgtaaccatt	ataagctgca	ataaaacaagt	tggggtgggc	8700
gaagaactcc	agcatgagat	ccccgcgctg	gaggatcatc	cagccggcgt	cccggaaaac	8760
gattccgaag	cccaaccttt	catagaaggc	ggcggtggaa	tcgaaatctc	gtagcacgtg	8820
tcagtccctg	tccctggcca	cgaagtgcac	g			8851

<210> 125

<211> 10474

<212> DNA

<213> Artificial Sequence

<220>

<223> p18genEPO Plasmid

<400> 125

cagttgcccgg	ccgggtcgcg	cagggcggaac	tccccccccc	acggctgctc	gccgatctcg	60
gtcatggccc	gcccggaggc	gtcccgggaag	ttcgtggaca	cgacctccga	ccactcgccg	120
tacagctcgt	ccaggcccg	cacccacacc	caggccagg	tggtgtccgg	caccacctgg	180
tccctggaccg	cgctgatgaa	cagggtcacg	tcgctccgga	ccacaccggc	gaagtgcgtc	240
tccacgaagt	cccgggagaa	cccagccgg	tcggtccaga	actcgaccgc	tccggcgacg	300
tgcgcgcgg	tgagcaccgg	aacggcactg	gtcaacttgg	ccatggatcc	agatttcgct	360
caagttagta	taaaaaagca	ggcttcaatc	ctgcagagaa	gcttgatata	gaattcctgc	420
agcccccg	atccgctcac	ggggacagcc	ccccccaaa	gccccagg	atgtaattac	480
gtccctcccc	cgctaggggg	cagcagcgag	ccgccccggg	ctccgctccg	gtccggcgct	540



ccccccgcat	ccccgagccg	gcagcgtgcg	gggacagccc	gggcacgggg	aaggtggcac	600
gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	cctgggggat	660
acggggcgcg	ggatccgctc	acggggacag	cccccccca	aagccccag	ggatgtaatt	720
acgtccctcc	cccgtaggg	ggcagcagcg	agccgcccg	ggctccgctc	cggcccgcg	780
ctccccccgc	atccccgagc	cggcagcgtg	cggggacagc	ccgggcacgg	ggaaggtggc	840
acgggatcgc	tttccctctga	acgttctctg	ctgctctttg	agcctgcaga	cacctggggg	900
atacgggggc	gcggatccgc	tcacggggac	agccccccc	caaagcccc	agggatgtaa	960
ttacgtccct	cccccgctag	ggggcagcag	cgagccgccc	ggggctccgc	tcgggtccgg	1020
cgctcccccc	gcataccccga	gccggcagcg	tgccggggaca	gcccgggcac	ggggaaggtg	1080
gcacgggatc	gctttctctct	gaacgcttct	cgctgctctt	tgagcctgca	gacacctggg	1140
ggatacgggg	ccgcggtacc	gctcacgggg	acagcccccc	cccaaagccc	ccagggatgt	1200
aattacgtcc	ctcccccgct	agggggcagc	agcgagccgc	ccggggctcc	gctccgggtc	1260
ggcgctcccc	ccgcatacccc	gagccggcag	cgctgcgggg	cagccccggg	acggggaagg	1320
tgccacggga	tcgctttcct	ctgaacgctt	ctcgctgctc	tttgagcctg	cagacacctg	1380
ggggatacgg	ggccgcggat	ccgctcacgg	ggacagcccc	cccccaaagc	ccccagggat	1440
gtaattacgt	ccctcccccg	ctagggggca	gcagcgagcc	gccccggggc	ccgctccggg	1500
ccggcgctcc	ccccgcctcc	ccgagccggc	agcgtgcggg	gacagcccg	gcacggggaa	1560
ggtggcacgg	gatcgctttc	ctctgaacgc	ttctcgctgc	tccttgagcc	tgacagacc	1620
tgggggatc	ggggccgcg	atccgctcac	ggggacagcc	ccccccaaa	gcccccagg	1680
atgtaattac	gctccctccc	cgctaggggg	cagcagcgag	ccgcccgggg	ctccgctccg	1740
gtccggcgct	ccccccgcat	ccccgagccg	gcagcgtgcg	gggacagccc	gggcacgggg	1800
aaggtggcac	gggatcgctt	tcctctgaac	gcttctcgct	gctctttgag	cctgcagaca	1860
cctgggggat	acggggcgga	ggatccacta	gttattaata	gtaatacaatt	acggggctcat	1920
tagttcatag	cccatatatg	gagttccgcg	ttacataact	tacggtaaat	ggcccgctcg	1980
gctgaccgcc	caacgacccc	cgcccattga	cgtaataaat	gacgtatgtt	cccatagtaa	2040
cgccaatagg	gactttccat	tgacgtcaat	gggtggacta	tttacggtaa	actgccact	2100
tgccagtaga	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	2160
aatggcccg	ctggcattat	gcccagtaga	tgaccttatg	ggactttcct	acttggcagt	2220
acatgacgt	attagtcate	gctattacca	tggttcgagg	tgagccccac	gttctgcttc	2280
actctcccca	tctccccccc	ctcccccccc	ccaattttgt	attttatttat	tttttaatta	2340
ttttgtgcag	cgatgggggg	gggggggggg	ggggcgcgcg	ccaggcgggg	cggggcgggg	2400
cgagggcgcg	ggcgggcgga	ggcgagagg	tgccggcgca	gccaatcaga	gcgggcgct	2460
ccgaaagtgt	ccctttatgg	cgaggcgggc	gcgccggcg	ccctataaaa	agcgaagcgc	2520
gcgccggggc	ggagtcgctg	cggttgcctc	gccccgtgcc	ccgctccgcg	ccgcctcgcg	2580
ccgcccgcgc	ggcctctgac	tgaccgcgtt	actcccacag	gtgagcgggc	gggacggccc	2640
ttctcctccg	ggctgtaatt	agcgcttggt	ttaatgacgg	ctcgtttctt	ttctgtggct	2700
gcgtgaaagc	cttaaagggc	tcggggagg	ccctttgtgc	gggggggagc	ggctcggggg	2760
gtgcgtgcgt	gtgtgtgtgc	gtggggagcg	ccgcgtgcgg	cccgcgctgc	ccggcgcgct	2820
tgagcgtgcg	gggggctttg	tgccgctccg	tgccgctccg	gtgtgcgcga	ggggagcgcg	2880
gccggggcg	gtgccccgcg	gtgccccggg	gctgcgaggg	gaacaaaggc	tgctgcgggg	2940
gtgtgtgcgt	gggggggtga	gcaggggggt	tgggcgcgcc	ggtcgggctg	taaccccccc	3000
ctgcaccccc	ctcccgaggt	tgctgagcac	ggcccggtct	cggggtgcgg	gctccgtgcg	3060
gggcgtggcg	cggggctcgc	cggtgcgggc	gggggggtgg	ggcaggtggg	ggtgcgggg	3120
ggggcggggg	cgccctcggg	cggggagggc	tcgggggagg	ggcgcgcgcg	ccccggagcg	3180
ccggcgcgct	tcgaggcgcg	gcgagccgca	gccattgect	tttatggtaa	tcgtgcgaga	3240
gggcgcagg	acttcccttg	tcccaaatct	ggcgagcccg	aaatctggga	ggcgccgccc	3300
cacccccctc	agcggggcg	ggcgaagcgg	tgccggcgcc	gcaggaagga	aatgggcggc	3360
gagggccttc	gtgcgtcgcc	gcgcgcgcgt	ccccctctcc	atctccagcc	tcggggctgc	3420
cgacggggga	cggtgccttt	cgggggggac	ggggcagggc	ggggttcggc	ttctggcggt	3480
tgaccggcg	ctctagatgc	atgctcgagc	ggccgcaggt	gtgatggata	tcctgcagaat	3540
tcgccccttc	tagaatgggg	gtgcacgggt	agtaactcg	ggctggggcg	tcggcgccgc	3600
ccgggtccct	gtttgagcgg	ggatttagcg	ccccggctat	tgccagggag	gtggctgggt	3660
tcaaggaccg	gcgacttgtc	aaggaccccc	gaagggggag	gggggtgggg	tgccctccac	3720
tgccagcggg	gacttggggg	agtccttggg	gatggcaaaa	acctgacctg	tgaaggggac	3780
acagtttggg	ggttgagggg	aagaagggtt	gggggttctg	ctgtgccagt	ggagagggaag	3840
ctgataagct	gataacctgg	gcgctggagc	caccacttat	ctgccagagg	ggaagcctct	3900
gtcacaccag	gattgaagtt	tgcccgagga	agtggatgct	ggtagctggg	gggggggtgt	3960
gcacacggca	gcaggattga	atgaaggcca	gggaggcagc	acctgagtg	ttgcatgggt	4020
ggggacagga	aggacagact	ggggcagaga	cgtggggatg	aagggaagct	tccttcacaa	4080
gcccaccttc	tcctccccc	cctgactctc	agcctggcta	tctgttctag	aatgtctcgc	4140
ctggctgtgg	cttctcctgt	ccctgctgtc	gctccctctg	ggcctccag	tcctgggcgc	4200
cccaccacgc	ctcatctgtg	acagccgagt	cctggagagg	tacctcttgg	aggccaagga	4260
ggccagaga	atcacgggtg	gaccccttcc	ccagcacatt	ccacagaact	cagctcagg	4320
gcttcaggga	actcctccca	gatccaggaa	cctggcactt	ggtttgggg	ggagtgggga	4380
agctagacac	tgccccctca	cataagaata	agtcctgggt	cccaaacca	tacctggaaa	4440
ctaggcaagg	agcaaagcca	gcagatccta	cgccctgtgg	gccagggcca	gagccttcag	4500
ggacccttga	ctccccgggc	tgtgtgcatt	tcagacgggc	tgtgctgaac	actgcagctt	4560



gaatgagaat	atcactgtcc	cagacaccaa	agttaatttc	tatgcctgga	agaggatgga	4620
ggtgagttcc	tttttttttt	tttttccctt	cttttggaga	atctcatttg	cgagcctgat	4680
tttggatgaa	aggggagaatg	atcgagggaa	aggtaaaatg	gagcagcaga	gatgaggctg	4740
cctgggcgca	gaggctcacg	tctataatcc	caggctgaga	tggccgagat	gggagaattg	4800
cttgagccct	ggagtttcag	accaacctag	gcagcatagt	gagatccccc	atctctacaa	4860
acatttaaaa	aaatttagtca	ggtgaagtgg	tgcatgggtg	tagtcccaga	tatttggaa	4920
gctgaggcgg	gaggatcgct	tgagcccagg	aatttgaggc	tgcatgagc	tgtgatcaca	4980
ccactgcact	ccagcctcag	tgacagagt	aggccctgtc	tcaaaaaaga	aaagaaaaaa	5040
gaaaaataat	gagggctgta	tggaatacat	tcatttattca	ttcactcact	cactcactca	5100
ttcattcatt	cattcattca	acaagtctta	ttgcatacct	tctgtttgct	cagcttggtg	5160
cttggggctg	ctgaggggca	ggagggagag	ggtgacatgg	gtcagctgac	tcccagagtc	5220
cactccctgt	aggctcgggca	gcaggccgta	gaagtctggc	agggccctggc	cctgctgtcg	5280
gaagctgtcc	tgccggggcca	ggccctgttg	gtcaactctt	cccagccgtg	ggagcccctg	5340
cagctgcctg	tggaataaagc	cgctcagtgcc	cttcgcagcc	tcaccactct	gcttcgggct	5400
ctgggagccc	aggtgagtag	gagcggacac	ttctgcttgc	cctttctgta	agaaggggag	5460
aagggctctg	ctaaggagta	caggaactgt	ccgtattcct	tccttttctg	tggcactgca	5520
gcgacctcct	gttttctcct	tgccagaagg	aagccatctc	ccctccagat	gcggcctcag	5580
ctgctccact	ccgaacaatc	actgctgaca	ctttccgcaa	actcttccga	gtctactcca	5640
atctcctccg	gggaaagctg	aagctgtaca	caggggaggg	ctgcaggaca	ggggacagat	5700
gacgtacaag	taagaattca	ctcctcaggt	gcaggctgcc	tatcagaagg	tgggtgctgg	5760
tgtggccaat	gccctggctc	acaaatacca	ctgagatctt	tttccctctg	ccaaaaatta	5820
tggggacatc	atgaagcccc	ttgagcatct	gacttctggc	taataaagga	aattttattt	5880
cattgcaata	gtgtgttggg	attttttggg	tctctcactc	ggaaggacat	atgggagctg	5940
aaatcattta	aaacatcaga	atgagtaatt	ggtttagagt	ttggcaacat	atgccatatt	6000
ctggctgcca	tgaacaaagg	tggtatataa	gaggtcatca	gtatatgaaa	cagccccctg	6060
ctgtccattc	cttattccat	agaaaagcct	tgacttgagg	ttagattttt	tttatatttt	6120
gttttgtgtt	atttttttct	ttaacatccc	taaaattttc	cttacatgtt	ttactagcca	6180
gatttttctc	cctctccctg	ctactcccag	tcatagctgt	ccctcttctc	ttatgaagat	6240
ccctcgacct	gcagcccaag	cttgcatgcc	tgacaggtcg	ctctagtggg	tccccgcccc	6300
cgtatccccc	aggtgtctgc	aggctcaaag	agcagcgaga	agcgttcaga	ggaaagcgat	6360
cccgtgccac	cttccccgtg	cccgggctgt	ccccgcacgc	tgccggctcg	gggatgcggg	6420
gggagcgccg	gaccggagcg	gagccccggg	cggtcgcctg	ctgcccccta	gcgggggagg	6480
gacgtaatta	catccctggg	ggctttgggg	gggggctgtc	cccgtgagcg	gatccgcggc	6540
cccgtatccc	ccaggtgtct	gcaggctcaa	agagcagcga	gaagcgttca	gaggaaaagc	6600
atcccgtgcc	accttccccg	tgccccggct	gtccccgcac	gctgccggct	cggggatgag	6660
gggggagcgc	cggaccggag	cggagccccg	ggcggctcgc	tgctgcccc	tagcggggga	6720
gggacgtaat	tacatccctg	ggggctttgg	gggggggctg	tccccgtgag	cggatccgcg	6780
gccccgtatc	cccaggtgtg	ctgcaggctc	aaagagcagc	gagaagcggt	cagaggaaag	6840
cgatcccgtg	ccaccttccc	cgtgccccgg	ctgtccccgc	acgctgccgg	ctcggggatg	6900
cggggggagc	gcccggaccg	agcggagccc	ggggcggtct	gctgtgtccc	cctagcgggg	6960
gagggacgta	attacatccc	tgggggcttt	gggggggggg	tgtccccgtg	agcgggatccg	7020
cggccccgta	tccccaggtg	gtctgcaggc	tcaaagagca	gcgagaagcg	ttcagaggaa	7080
agcgatcccc	tgccaccttc	cccgtgcccc	ggctgtcccc	gcacgctgcc	ggctcggggg	7140
tgcgggggga	gcgccggacc	ggagcggagc	cccggcgggc	tcgctgctgc	ccctagcggg	7200
gggaggggac	taattacatc	cctgggggct	ttgggggggg	gctgtccccg	tgagcggatc	7260
cgcgggcccc	tatccccag	gtgtctgcag	gctcaaagag	cagcgagaag	cgttcagagg	7320
aaagcgatcc	cgtgccacct	tccccgtgcc	cgggctgtcc	ccgcacgctg	ccggctcggg	7380
gatgcggggg	gagcgccgga	ccggagcgga	gccccggggc	gctcgctgct	gccccctagc	7440
gggggagggg	cgtaattaca	tccctggggg	ctttgggggg	gggctgtccc	cgtgagcgga	7500
tccgcggccc	cgtatcccc	aggtgtctgc	aggtcaaag	agcagcgaga	agcgttcaga	7560
ggaaagcgat	cccggtgccac	cttccccgtg	cccgggctgt	ccccgcacgc	tgccggctcg	7620
gggatgcggg	gggagcgccg	gaccggagcg	gagccccggg	cggctcgctg	ctgcccccta	7680
gcgggggagg	gacgtaatta	catccctggg	ggctttgggg	gggggctgtc	cccgtgagcg	7740
gatccgcggg	gctgcaggaa	ttcgtaatac	tggtcatagc	tggttctctg	gtgaaattgt	7800
tatccgctca	caattccaca	caacatacga	gccggaagca	taaagtgtaa	agcctggggg	7860
gcctaattgag	tgagctaact	cacattaatt	cgttgccgct	cactgcccgc	tttccagctg	7920
ggaaacctgt	cgtgccagct	gcattaatga	atcggccaa	gcgcggggag	aggcgggttg	7980
cgtattgggg	gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	cgttcgggctg	8040
cggcgagcgg	tatacagctc	ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	8100
aacgcaggaa	agcaaaaggc	agcaaaaggc	cagcaaaagg	ccaggaacgc	taaaaaggcc	8160
gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacia	aaatcgacgc	8220
tcaagtacga	ggtggcgaaa	cccgacagga	ctataaagat	accaggcggt	tccccctgga	8280
agctccctcg	tgctccgacc	tgctccgacc	ctgcgcttta	coggataacct	gtccgctctt	8340
ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	cagttcggtg	8400
taggtcggtc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	8460
gccttatccg	gttaactatcg	tcttgagctc	aacccggtaa	gacacgactt	atcgccactg	8520
gcagcagcca	ctggtaacag	gatttagcaga	gcgaggtatg	tagggcggtg	tacagagttc	8580

```
ttgaagtgggt ggcctaacta cggctacact agaaggacag tatttggtat ctgcgctctg 8640
ctgaagccag ttaccttcgg aaaaagaggt ggttagctctt gatccggcaa acaaaccacc 8700
gctggtagcg gtgggtttttt tggttgcaag cagcagatta cgcgagaaa aaaaggatct 8760
caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaaacga aaactcacgt 8820
taagggatgt ttggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa 8880
aaatgaagtt ttaaataaat cttaaagtata tatgagtaaa cttgggtctga cagttaccaaa 8940
tgcttaatac gtgaggcacc tatctcagcg atctgtctat ttcgttcac ccatagttgcc 9000
tgactccccg tcgtgtagat aactacgata cgggaggggt taccatctgg cccagtgct 9060
gcaatgatac cgcgagaccc acgctcacgg gctccagatt tatcagcaat aaaccagcca 9120
gccggaaggg ccgagcgcag aagtggctct gcaactttat ccgcctccat ccagtctatt 9180
aattgttgcc ggaagctag agtaagtagt tcgcagttta atagtttgcg caacgttggt 9240
gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc 9300
gggtcccaac gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcgggttagc 9360
tccttcggtc ctccgatcgt tgtcagaagt aagttggccg cagtgttatc actcatggtt 9420
atggcagcac tgcataattc tcttactgtc atgccatccg taagatgctt tctgtgtagt 9480
ggtgagtact caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc 9540
ccggcgtaaa taccgagtaa taaccgagaa catagcagaa ctttaaaagt gtcacatt 9600
ggaaaacggt cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg 9660
atgtaaccga ctctgacacc caactgatct tcagcatctt ttactttcac cagcgtttct 9720
gggtgagcaa aaacaggaag gcaaaaagg gcaaaaagg gaataaggcg gacacggaaa 9780
tgttgaatac tcatactctt cctttttcaa tattattgaa gcatttatca gggttattgt 9840
ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc 9900
acatttcccc gaaaagtgcc acctgacgta gtttaacaaa aaaagccgc cgaagcgggc 9960
tttattacca agcgaagcgc cattcgccat tcaggctgcg caactgttgg gaagggcgat 10020
cgggtcgggg ctctctcgta ttacgccagc tggcgaaaagg gggatgtgct gcaagggcat 10080
taagttaggt aacgccaggg ttttccaggt cagcagcttg taaaacgacg gccagtcctg 10140
aatcagactc acttaaggcc ttgactagag ggtcgacggg atacagacat gataagatac 10200
attgtagat ttggacaaac cacaactaga atgacagtgaa aaaaatgctt tatttgtgaa 10260
atttgtgatg atttgtctt attataagct atataagct gcaataaaca agttggggtg 10320
ggcgaagaac tccagcatga gatccccgcg ctggaggatc atccagccgg cgtcccgga 10380
aacgattccg aagcccaacc tttcatagaa ggcggcggtg gaatcgaaat ctcgtagcac 10440
gtgtcagtc tgctcctcg ccacgaagtg cagc 10474
```

<210> 126

<211> 6119

<212> DNA

<213> Artificial Sequence

<220>

<223> p18attBZeoeGFP Plasmid

<400> 126

```
cagttgcccgg ccgggtcgcg cagggcgaa ccccccccc acggtgctc gccgatctcg 60
gtcatggccg gcccgaggc gtcccggaag ttcgtggaca cgacctcga ccactcggcg 120
tacagctcgt ccaggccgcg caccacacac caggcaggg tggtgtccgg caccacctgg 180
tcctggaccg cgtgatgaa cagggtcacg tcgtcccggg ccacaccggc gaagtgcgtc 240
tccacgaagt cccgggagaa cccgagccgg tcggtccaga actcgaccgc tccggcgacg 300
tcgcgcgcgg tgagcaccgg aacggcactg gtcaacttgg ccatggatcc agatttcgct 360
caagttagta taaaaaagca ggcttcaatc ctgcagagaa gcttgggctg caggtcgagg 420
gatcttcata agagaagagg gacagctatg actgggagta gtcaggagag gaggaaaaat 480
ctggctagta aaacatgtaa ggaaaatttt agggatgta aagaaaaaaa taacacaaaa 540
caaaatataa aaaaaatcta acctcaagtc aaggcttttc tatggaataa ggaatggaca 600
gcaggggggt gtttcatata ctgatgacct ctttatagcc acctttgttc atggcagcca 660
gcatatggca tatgttgcca aactctaaac caaatactca ttctgatgtt ttaaattgatt 720
tgccctccca tatgtccttc cgagtggagag acacaaaaaa ttccaacaca ctattgcaat 780
gaaaataaat ttcctttatt agccagaagt cagatgctca aggggcttca tgatgtcccc 840
ataatttttg gcagagggaa aaagatctca gtggattttg tgagccaggg cattggccac 900
accagccacc accttctgat aggcagcctg cacctgagga gtgaattctt acttgtagag 960
ctcgtccatg ccgagagtga tcccggcgcc ggtcacgaac tccagcagga ccatgtgatc 1020
gcgcttctcg ttgggtctct tgctcagggc ggactgggtg ctcaggtagt ggttgcggg 1080
cagcagcacg gggccgtcgc cgatgggggt gttctgctgg tagtggctcg ttgatgccgt tcttctgctt 1200
gctgccgtcc tcgatgttgt tgtggctgtt gttctgctgg gaagttcacc tccagcttgt gccccaggat 1260
gtcggccatg atatagacgt cgtggccttt cagctcgatg cggttcacca ggggtgtcgcc 1320
gttgccgtcc acctcggcgc ggttcttgta gttgccgtcg tccttgaaga agatgggtgc 1380
ctcctggacg tagccttcgg gcatggcgga cttgaagaag tctgtgctgt ccatgtggtc 1440
ggggtagcgg ctgaagcact gcacgccgta ggtcagggtg gtcacgaggg tgggcccagg 1500
```

cacgggacgc	ttgcccgggtg	tgcagatgaa	cttcagggtc	agcttgccgt	aggtggcatc	1560
gccctcgccc	tgcgcccga	cgctgaactt	gtggccggtt	acgtcgccgt	ccagctcgac	1620
caggatgggc	accaccccgg	tgaacagctc	ctcgcccttg	ctcaccatgg	tggcgaattc	1680
tttgcacaaa	tgatgagaca	gcacaacaac	cagcacgttg	cccaggagct	gtaggaaaaa	1740
gaagaaggca	tgaacatggt	tagcagaggc	tctagagccg	ccggtcacac	gccagaagcc	1800
gaaccccgc	ctgccccgct	ccccccgaag	gcagccgctc	ccctgcccga	gccccgaggc	1860
tgagatgga	gaaggggacg	gcccgcgggc	gacgcacgaa	ggccctcccc	gcccatttcc	1920
ttcctgccc	cgccgcaccg	cttcgcccgc	gcccgcctaga	gggggtgccc	cgccgcctcc	1980
cagatttcgg	ctccgcccga	tttgggacaa	aggaagtccc	tgcgcccctc	cgcacgatta	2040
ccataaaaagg	caatggctgc	ggctcgccgc	gcctcgacag	ccgcccggcg	tccggggccc	2100
ccgccccct	cccccgagcc	ctccccggcc	cgaggcgccc	ccgccccgcc	cggcaccccc	2160
acctgcccgc	accccccgcc	cggcacggcg	agcccccgcc	cacgccccgc	acggagcccc	2220
gcacccgaag	ccgggcccgt	ctcagcaact	cggggagggg	gggtgcaggg	gggggttacag	2280
cccgcacg	gcccgcacac	ccccctgctc	ccccccacg	cacacacccc	gcacgcagcc	2340
tttgttcccc	tgcgagcccc	cccgccacgc	ggggcacccg	ccccggccgc	gctccccctc	2400
cgccacgcgc	gagcgcacaa	agccccgcgc	cacacacacg	gcccctcacag	ccgcccgggc	2460
gcccggggcg	cacgcggcg	tccccacgca	ctttaaggct	cacgcacccc	ccgagccgct	2520
cccccccgc	caaagggccc	tcccggagcc	tacagcccgg	ttcacgcagc	cacagaaaaa	2580
aaacgagccg	tcattaaacc	aagcgttaac	agagccgggg	aggagaaggg	ccgtccccgc	2640
cgctcacctg	tgggagttaac	ggcgtcagtc	agagccgggg	cgggcgccgc	gaggcgccgc	2700
ggagcggggc	acggggcgaa	ggcaacgcag	cgactcccgc	ccgcccgcgc	cttcgctttt	2760
tatagggcgc	ccgcccgcgc	cgccctcgca	taaaaggaaa	ctttcgggagc	gcccgcctct	2820
gattggctgc	cgccgcaccc	ctccgctcgc	ccccgcccc	ccccctcgcc	cgccccgccc	2880
cgccctggcg	gcccgcaccc	cccccccgcc	cccatcgctg	cacaaaaata	ttaaaaataa	2940
aataaatata	aaattggggg	tggggagggg	ggggagatgg	ggagagtga	gcagaacgtg	3000
cggtcacctg	cgacccatgg	taatagcgat	gactaataac	tagatgtact	gccaagttag	3060
aaagtcccat	aaggctcatgt	actgggcata	atgccaggcg	ggccatttac	cgctcattgac	3120
gtcaatatgg	ggcgtacttg	gcataatgata	cacttgatgt	actgccaagt	gggcagttaa	3180
cgttaaatag	tccacccatt	gacgtcaatg	gaaagtcctc	attggcggtt	ctatgggaac	3240
atacgtcatt	attgacgtca	atgggcccgg	gtcgttgggc	ggtcagccag	gcccggccatt	3300
taccgtaagt	tatgtaacgc	ggaactccat	atatgggcta	tgaactaatg	accccgtaac	3360
tgattactat	taataactag	aggatccccg	ggtagccgag	tccaattcgt	aatcatggtc	3420
atagctgttt	cctgtgtgaa	attgttatcc	gctcacaatt	ccacacaaca	tacgagccgg	3480
aagcataaag	tgttaaagct	ggggtgccta	atgagtgcgc	taactcacat	taattgcgtt	3540
gcgtcacctg	cccgccttcc	agtcgggaaa	ctcgtcgctg	cagctgcatt	aatgaatcgg	3600
ccaacgcgcg	gggagaggcg	gtttgcgtat	tgggcgctct	tccgcttcc	cgctcactga	3660
ctcgctcgcc	tccgtcgctc	ggctgcggcg	agccggtatc	gctcactcaa	aggcggtaat	3720
acgggttatcc	acagccatcag	gggataaacg	aggaagaaac	atgtgagcaa	aaggccagca	3780
aaaggccagg	aaccgtaaaa	aggccgcgct	gctggcgctt	ttccataggc	tccgcccccc	3840
gcagagcatt	cacaaaaaat	gacgctcaag	tcagagggtg	cgaaaccoga	caggactata	3900
aagataccag	cggtttcccc	ctgggaagct	cctcgtcgcg	tctcctgttc	cgaccctgcc	3960
gcttaccgga	tacctgtccg	cctttctccc	ttcggaagc	gtggcgcttt	ctcatagctc	4020
acgctgtagg	tatctcagtt	cggtgtaggt	cgttcgctcc	aagctgggct	gtgtgcacga	4080
accccccggt	cagcccgacc	gctgcgcctt	atccggtaac	tatcgtcttg	agtcacaccc	4140
ggtaagacac	gacttatcgc	cactggcagc	agccactgg	aacaggatta	gcagagcgag	4200
gatagttagc	gggtgctacg	agttcttgaa	gtgggtggc	aactacggct	acactagaag	4260
gcagtatatt	gggtatctgc	ctctgctgaa	gccagttacc	ttcggaaaaa	gagttggtag	4320
ctcttgatcc	ggcaaacaaa	ccaccgctgg	tagcgggtgg	ttttttgttt	gcaagcagca	4380
gattacgcgc	agaaaaaaag	gatctcaaga	agatcccttg	atctttttct	cggggtctga	4440
cgctcagtg	aacgaaaaac	cacgttaagg	gatttttggt	atgagattat	caaaaaggat	4500
cttcacctag	atccttttaa	attaaaaatg	aagtttttaa	tcaatctaaa	gtatatatga	4560
gtaaaacttg	cttgacagtt	accaatgctt	aatcagtgag	gcacctatct	cagcgatctg	4620
tctatttcgt	tcatccatag	ttgcctgact	ccccgctg	tagataacta	cgatacggga	4680
gggcttacca	tctggcccca	gtgctgcaat	gataccgcga	gacccacgct	caccggctcc	4740
agatttatca	gcaataaaac	agccagccgg	aagggccgag	cgcagaagtg	gtcctgcaac	4800
tttatccg	tcacccag	ctattaattg	ttgcccggaa	gctagagtaa	gtagttcgcc	4860
agtttaata	ttgcgcaacg	ttgttgccat	tgctacaggg	atcgtgggtg	cacgctcgct	4920
gtttgggtat	gcttcattca	gctccgggtc	ccaacgatca	aggcgagtta	catgatcccc	4980
catgttgtgc	aaaaaagcgg	ttagttcctt	cggtcctccg	atcgttgtca	gaagtaagtt	5040
ggccgcagtg	ttatcactca	tggttatggc	agcactgcat	aattctctta	ctgtcatgcc	5100
atccgtaaga	tgcttttctg	tgactgggtg	gtactcaacc	aagtcattct	gagaatagtg	5160
tatgcggcga	ccgagttgct	cttgccggcg	gtcaatacgg	gataataccg	cgccacatag	5220
cagaacttta	aaagtgtcta	tcattggaaa	acgttcttcg	gggcgaaaaa	tctcaaggat	5280
cttaccgctg	ttgagatcca	gttcgatgta	accactcgt	gcacccaact	gatcttcagc	5340
atcttttact	ttcaccagcg	tttctgggtg	agcaaaaaaa	ggaaggcaaa	atgcccgaaa	5400
aaaggggaata	agggcgacac	ggaaaatgtg	aatactcata	ctcttccttt	ttcaatatta	5460
ttgaagcatt	tatcagggtt	attgtctcat	gagcggatac	atatttgaat	gtatttagaa	5520

aaataaaca	ataggggttc	cgcgacacatt	tccccgaaaa	gtgccacctg	acgtagttaa	5580
caaaaaaaag	cccgcgaag	cggtctttat	taccaagcga	agcgccattc	gccattcagg	5640
ctgcgcaact	gttgggaagg	gcgatcgggtg	cgggcctctt	cgctattacg	ccagctggcg	5700
aaagggggat	gtgctgcaag	gcgattaagt	tgggtaacgc	caggggtttc	ccagtcacga	5760
cggtgtaaaa	cgacggccag	tccgtaatac	gactcactta	aggccttgac	tagaggggtc	5820
acgggtatata	gacatgataa	gatacattga	tgagtttgga	caaaccacaa	ctagaatgca	5880
gtgaaaaaaa	tgctttat	gtgaaatttg	tgatgctatt	gctttatttg	taaccattat	5940
aagctgcaat	aaacaagttg	gggtgggcga	agaactccag	catgagatcc	ccgcgctgga	6000
ggatcatcca	gccggcgctc	cggaaaacga	ttccgaagcc	caacctttca	tagaaggcgg	6060
cgggtggaatc	gaaatctcgt	agcacgtgtc	agtctgtctc	ctcggccacg	aagtgcacg	6119

<210> 127

<211> 5855

<212> DNA

<213> Artificial Sequence

<220>

<223> pCXLamInt Plasmid (Wildtype Integrase)

<400> 127

gtcgacattg	attattgact	agttattaat	agtaatcaat	tacgggggtca	ttagttcata	60
gcccataat	ggagttccgc	gttacataac	ttacggtaaa	tggcccgcc	ggctgaccgc	120
ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	acgccaatag	180
ggactttcca	ttgacgtcaa	tgggtggact	atttacggta	aactgcccac	ttggcagtag	240
atcaagtgt	tcatatgcca	agtacgcccc	ctattgacgt	caatgacggg	aaatggcccg	300
cctggcatta	tgcccagtag	atgaccttat	gggactttcc	tacttggcag	tacatctacg	360
tattagtc	cgctattacc	atgggtcgag	gtgagcccca	cgttctgctt	cactctcccc	420
atctccccc	cctccccacc	cccaattttg	tatttattta	ttttttaatt	attttgtgca	480
gcgatggggg	cggggggggg	gggggcccgc	gccaggcggg	gcggggcggg	gcgagggggc	540
gggcggggcg	aggcggagag	gtgcggcggc	agccaatcag	agcggcgccg	tccgaaagtt	600
tccttttatg	gcgaggcggc	ggcgggcggc	gccctataaa	aagcgaagcg	cgcgggcggg	660
gggagtcgct	gcgttgccct	cgcccgcgtc	cccgcctcgc	gcccgcctcg	gcccgcggcc	720
ccggctctga	ctgaccgcgt	tactcccaca	ggtagcgggg	cgggacggcc	cttctccctc	780
gggctgtaat	tagcgcttgg	tttaatgacg	gctcgtttct	tttctgtggc	tgcgtagaag	840
ccttaaaagg	ctccgggagg	gccctttgtg	cgggggggag	cggtcggggg	gggtgcgtgc	900
tgtgtgtgtg	cgtggggagc	gccgcgtgcg	gcccgcgctg	cccgccggct	gtgagcgctg	960
cgggcgcggc	gcggggcctt	gtgcgctccg	cgtgtgcgcg	aggggagcgc	ggccgggggg	1020
ggtgccccgc	gggtgcgggg	ggctgcgagg	ggaacaaaag	ctgcgtgcgg	gggtgtgtgc	1080
tgggggggtg	agcagggggg	gtggggcgcg	cggtcgggct	gtaaccccc	cctgcacccc	1140
cctccccgag	ttgctgagca	cgcccgggct	tcgggtgcgg	ggctccgtgc	ggggcggtgg	1200
gcggggctcg	ccgtgcgggg	cggggggttg	cggcagggtg	gggtgcgggg	cggggcgggg	1260
ccgcttcggg	ccggggaggg	ctcgggggag	gggcgcggcg	gccccggagc	gcccggcggt	1320
gtcgaggcgc	ggcgagccgc	agccattgcc	ttttattgta	atcgtgcgag	agggcgcgag	1380
gacttccttt	gtcccaaatc	tggcggagcc	gaaatctggg	aggcgcggcc	gcacccccct	1440
tagcggcgcg	gggcgaagcg	gtgcggcgcc	ggcaggaagg	aaatgggcgg	ggagggccct	1500
cgtgcgtcgc	cgcgcgcgcg	tccccttctc	catctccagc	ctcggggctg	ccgcaggggg	1560
acggctgcct	tcggggggga	cggggcaggg	cggggttcgg	cttctggcgt	gtgacggcg	1620
gctctagagc	ctctgctaac	catgttcatg	ccttcttctt	tttctacacg	ctcctgggca	1680
acgtgctggt	tggtgtgctg	tctcatcatt	ttggcaaga	attcatggga	agaaggcgaa	1740
gtcatgagcg	ccgggattta	ccccctaacc	tttatataag	aaacaatgga	tattactgct	1800
acagggaccc	aaggacgggt	aaagagtttg	gattagggcag	agacaggcga	atcgcaatca	1860
ctgaagctat	acaggccaac	attgagttat	tttcaggaca	caaacacaag	cctctgacag	1920
cgagaatcaa	cagtataat	tccgttacgt	tacattcatg	gcttgatcgc	tacgaaaaaa	1980
tcctggccag	cagaggaatc	aagcagaaga	cactcataaa	ttacatgagc	aaaattaaag	2040
caataaggag	gggtctgcct	gatgctccac	ttgaagacat	caccacaaaa	gaaattgctg	2100
caatgctcaa	tggatacata	gacgagggca	aggcggcgct	agccaagtta	atcagatcaa	2160
cactgagcga	tgcatccga	gaggcaatag	ctgaaggcca	tataacaaca	aacctgtctg	2220
ctgccactcg	cgcagcaaaa	tcagagggtaa	ggagatcaag	acttacggct	gacgaatacc	2280
tgaaaattta	tcaagcagca	gaatcatcac	catgttggtc	cagacttgca	atgggaactg	2340
ctgttggtac	cgggcaacga	gttggtgatt	tatgcgaaat	gaagtggctc	gatatcgtag	2400
atggatatct	ttatgtcgag	caaagcaaaa	caggcgtaaa	aattgccatc	ccaacagcat	2460
tgcatattga	tgctctcgga	atatcaatga	aggaacacat	tgataaatgc	aaagagattc	2520
ttggcggaga	aaccataatt	gcactctact	gtcgcgaacc	gctttcatcc	ggcacagtat	2580
caagggtattt	tatgcgcgca	cgaaaagcat	caggctcttc	cttcgaaggg	gatccgccta	2640
cctttcacga	gttctgcagt	ttgtctgcaa	gactctatga	gaagcagata	agcgataagt	2700
ttgctcaaca	tcttctcgga	cataagtcgg	acattctggc	atcacagtat	cgtgatgaca	2760
gaggcagggg	gtgggacaaa	attgaaatca	aataagaatt	cactcctcag	gtgcaggctg	2820

cctatcagaa	ggtggtggct	ggtgtggcca	atgccctggc	tcacaaatac	cactgagatc	2880
ttttccctc	tgccaaaaat	tatggggaca	tcatagaagcc	ccttgagcat	ctgacttctg	2940
gctaataaag	gaaattttatt	ttcattgcaa	tagtgtgttg	gaattttttg	tgtctctcac	3000
tcggaaggac	atatgggagg	gcaaatcatt	taaaacatca	gaatgagtat	ttggttttaga	3060
gtttggcaac	atatgccata	tgctggctgc	catgaacaaa	ggtgggtata	aagaggtcat	3120
cagtatatga	aacagccccc	tgctgtccat	tccttattcc	atagaaaagc	cttgacttga	3180
ggttagattt	tttttatatt	ttgttttgtg	ttattttttt	ctttaacatc	cctaaaattt	3240
tccttacatg	ttttactagc	cagatttttc	ctcctctcct	gactactccc	agtcatagtc	3300
gtccctcttc	tcttatgaag	atccctcgac	ctgcagccca	agcttggcgt	aatcatggtc	3360
atagctgttt	cctgtgtgaa	attgttatcc	gctcacaatt	ccacacaaca	tacgagccgg	3420
aagcataaag	tgtaaagcct	ggggtgccta	atgagtggagc	taactcacat	taattgcgtt	3480
gcgctcactg	cccgccttcc	agtcgggaaa	cctgtcgtgc	cagcggatcc	gcactcfaat	3540
tagtcagcaa	ccctagtccc	gccccatacc	ccgcccatacc	cgccccatac	tccgccaggt	3600
tccgccattt	ctccgccccca	tggtgacta	atttttttta	tttatgcaga	ggccgaggcc	3660
gcctcggcct	ctgagctatt	ccagaagtag	tgaggaggct	tttttggagg	cctaggcttt	3720
tgcaaaaagc	taacttgttt	attgcagctt	ataatgggta	caaataaagc	aatagcatca	3780
caaatttcac	aaataaagca	tttttttcac	tgcatctctag	ttgtgggttg	tccaaactca	3840
tcaatgtatc	ttatcatgtc	tggtatccgc	gcattaatga	atcggccaac	gcgcggggag	3900
agggcggttt	cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	3960
cggtcggctg	cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	4020
atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaggc	cagcaaaagg	ccaggaaccg	4080
taaaaaggcc	gcgttgctgg	cggttttcca	taggtccgcg	ccccctgacg	agcatcacaa	4140
aaatcgacgc	tcaagtcaga	ggtggcgaaa	cccgacagga	ctataaagat	accaggcggt	4200
tccccctgga	agctccctcg	tgcgctctcc	tggtccgacc	ctgcccgtta	ccggatacct	4260
gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcaa	tgctcacgct	gtaggtatct	4320
cagttcgggtg	taggtcggtc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	4380
cgaccgctgc	gccttatccg	gtaactatcg	tcttgagtcg	aaccgggtaa	gacacgactt	4440
atcgccactg	gcagcagcca	ctggtaacag	cttagcaga	gcgaggtatg	taggcggtgc	4500
tacagagttc	ttgaagtggg	ggcctaacta	cggctacact	agaaggacag	tatttggtat	4560
ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	4620
acaaaccacc	gctggtagcg	gtgggttttt	tgtttgcaag	cagcagatta	cgcgagaaaa	4680
aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaaacga	4740
aaactcacgt	taagggattt	tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	4800
tttaaattaa	aaatgaagtt	ttaaatcaat	ctaaagtata	tatgagtaaa	cttgggtctga	4860
cagttaccaa	tgcttaatca	gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcatc	4920
catagttgcc	tgactccccg	tcgtgttagt	aactacgata	cgggaggggt	taccatctgg	4980
ccccagtgtc	gcaatgatac	cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	5040
aaaccagcca	gccggaaggg	ccgagcgcag	aagtggtcct	gcaactttat	ccgcctccat	5100
ccagtctatt	aattgtttgcc	gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	5160
caacgttgtt	gccattgcta	caggcatcgt	ggtgtcacgc	tcgtcggttg	gtatggcttc	5220
attcagctcc	ggttcccaac	gatcaaggcg	agttacatga	tcccccatgt	tgtgcaaaaa	5280
agcggttagc	tccttcgggtc	ctccgatcgt	tgtcagaagt	aagttggccg	cagtggttatc	5340
actcatgggt	atggcagcac	tgcataattc	tcttactgtc	atgccatccg	taagatcctt	5400
ttctgtgact	ggtgagtact	caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	5460
ttgtctctgc	ccggcgctcaa	tacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	5520
gctcatcatt	ggaaaaagct	cttcggggcg	aaaactctca	aggatcttac	cgctgttgag	5580
atccagttcg	atgtaaccca	ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	5640
cagcgtttct	gggtgagcaa	aaacaggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	5700
gacacggaaa	tggtgaatac	tcatactctt	cctttttcaa	tattattgaa	gcatttatca	5760
gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	5820
ggttccgcgc	acatttcccc	gaaaagtgcc	acctg			5855

<210> 128  
 <211> 303  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Human FER-1 Promoter

<400> 128						
tccatgacaa	agcactttttt	gagcccaagc	ccagcctagc	tcgagctaaa	cgggcacaga	60
gacgccaccg	ctgtcccaga	ggcagtcggc	taccgggtccc	cgctcccagag	ctccgccaga	120
gcgcgcgagg	gcctccagcg	gcccgccttc	ccccacagca	ggggcggggt	cccgcgccca	180
ccggaaggag	cgggctcggg	gcgggcggcg	ctgattggcc	ggggcgggcc	tgacgccgac	240
gcggctataa	gagaccacaa	gcgacccgca	gggcccagacg	ttcttcgccg	agagtcgggt	300
acc						303

<210> 129  
<211> 6521  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> pIRES-BSR Plasmid

<400> 129  
tcaatattgg ccattagcca tattattcat tgggtatata gcataaatca atattgggcta 60  
ttggccattg catacgttgt atctatatca taatatgtac atttatattg gctcatgtcc 120  
aatatgaccg ccatgttggc attgattatt gactagttat taatagtaat caattacggg 180  
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc 240  
gcctggctga ccgcccacg acccccgccc attgacgtca ataatgacgt atgttcccat 300  
agtaacgcca atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaactgc 360  
ccacttggca gtacatcaag tgtatcatat gccaaagtccg cccctattg acgtcaatga 420  
cggtaaatgg ccgcctggc attatgccc a gtacatgacc ttacgggact ttctacttg 480  
gcagtacatc tacgtattag tcatcgctat taccatgggtg atgcggtttt ggcagtacac 540  
caatggggcgt ggatagcggg ttgactcacg gggatttcca agtctccacc ccattgacgt 600  
caatggggagt ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaacaactg 660  
cgatcgcccg ccccggttgac gcaaatgggc ggttaggcgtg tacgggtggga ggtctatata 720  
agcagagctc gtttagtgaa ccgtcagatc actagaagct ttattgcggt agtttatcac 780  
agttaaattg ctaacgcagt cagtgcctct gacacaacag tctcgaactt aagctgcagt 840  
gactctctta aggtagcctt gcagaagtgt gtcgtgaggg actgggcagg taagtatcaa 900  
ggttacaaga caggtttaag gagaccaata gaaactgggc ttgtcgagac agagaagact 960  
cttgcggtttc tgataggcac ctattgggtct tactgacatc cactttgcct ttctctccac 1020  
aggtgtccac tcccagttca attacagctc ttaaggctag agtacttaac acgactcact 1080  
ataggctagc tcttcttgac caccgcgtcga gcctgcctct agggcgccca attccgcccc 1140  
tctccctccc cccccctaa cgttactggc cgaagccgct tgggaataagg ccggtgtgag 1200  
tttgtctata tgtgattttc caccatattg ccgtctttt gcaatgtgag gggccggaaa 1260  
cctggccctg tcttcttgac gagcattcct aggggtcttt aagcttcttg aagacaaaca 1320  
caaggctctg tgaatgtcgt gaaggaagca gttcctcttg aagcttcttg aagacaaaca 1380  
acgtctgtag cgaccctttg caggcagcgg aacccccac ctggcgacag gtgcctctgc 1440  
ggccaaaagc cagctgtata agatacacct gcaaaaggcg cacaaccca gtgcacgtct 1500  
gtgagttgga tagttgtgga aagagtcaaa tggctctcct caagcgtatt caacaagggg 1560  
ctgaaggatg ccagaaagg accccattgt atgggatctg atctggggcc tcggtgcaca 1620  
tgctttacat gtgttttagtc gaggttaaaa aaacgtctag gcccccgaa ccacggggac 1680  
ttgtttttcc ttgaaaaac acgatgataa gcttgccaca acccaccatg aaaacattta 1740  
acattttctc acaagatcta gaattagtag aagtgcgac agagaagatt acaatgcttt 1800  
atgaggataa taaacatcat gtgggagcgg caattcgtac gaaaacagga gaaatcatt 1860  
cgcgagtaca tattagagcg tatataggac gagtaaacgt ttgtgcagaa gccattgcga 1920  
ttggtagtgc agtttcgaat ggacaaaagg attttgacac gattgtagct gttagacacc 1980  
cttattctga cgaagtagat agaagtattc gagtggtaag tccttgtggt atgtgtaggg 2040  
agttgatttc agactatgca ccagattgtt ttgtgttaat agaaatgaat ggcaagttag 2100  
tcaaaactac gattgaagaa ctcatccac tcaaatatac ccgaaattaa aagttttacc 2160  
ataccaagct tggcgggcgg ccgcttccct ttagtgaggg ttaatgcttc gagcagacat 2220  
gataagatac attgatgagt ttggacaaac cacaactaga atgcagtga aaaaatgctt 2280  
tatttgtgaa atttgtgatg ctattgcttt atttghtaacc attataagct gcaataaaca 2340  
agttaacaac aacaattgca ttcatthttat gtttcagggt cagggggaga tgtgggaggt 2400  
tttttaaaagc aagtaaaaacc tctacaaatg ttgtaaaaatc cgataaggat cgatccgggc 2460  
tggcgtaata gcgaagaggg ccgcaccgat cgcccttccc aacagttgcg cagcctgaat 2520  
ggcgaatgga cgcgcctgt agcggcgcac taagcgcggc ggggtgtggtg gttacgcgca 2580  
gcgtgaccgc tacacttgcc agcgccttag cgcccgctcc ttctcgctttc ttcccttct 2640  
ttctcgccac gttcgccggc ttccccgctc aagctctaaa tcgggggctc cctttagggt 2700  
tccgatttag agctttacgg cacttcgacc gcaaaaaact tgatttgggt gatggttcac 2760  
gtagtgggce atcgccctga tagacggttt ttcgcccttt gacgttggag tccagttct 2820  
ttaatagtgg actcttgttc caaactggaa caaactcaa ccctatctcg gtctattctt 2880  
ttgatttata agggattttg ccgatttcgg cctattgggt aaaaaatgag ctgatttaac 2940  
aaatatttaa cgcgaatttt aacaaatata taacgtttac aatttcgct gatcggtat 3000  
tttctcctta cgcactctgt cggtatttca caccgcatac gcggatctgc gcagcaccat 3060  
ggcctgaaat aacctctgaa agaggaactt ggttaggtac cttctgagggc ggaagaacc 3120  
agctgtggaa ttgtgtcag ttagggtgtg gaaagtcgcc aggcctccca gcaggcagaa 3180  
gtatgcaaag catgcatctc aattagtcag caaccaggtg tggaaagtcc ccaggctccc 3240  
cagcaggcag aagtatgcaa agcatgcac tcaattagtc agcaaccata gtcccgccc 3300  
taactccgcc catcccgcc ctaactccgc ccagttccgc ccattctcg cccatggct 3360  
gactaatttt ttttatttat gcagaggccg aggcgcctc ggcctctgag ctattccaga 3420  
agtagtgagg aggtttttt ggaggcctag gcttttgcaa aaagcttgat tcttctgaca 3480

caacagtctc	gaacttaagg	ctagagccac	catgattgaa	caagatggat	tgcacgcagg	3540
ttctccggcc	gcttgggtgg	agaggctatt	cggctatgac	tgggcacaac	agacaatcgg	3600
ctgctctgat	gccgccgtgt	tccggctgtc	agcgcagggg	cgcccggttc	tttttgtcaa	3660
gaccgacctg	tccgggtgcc	tgaatgaact	gcaggacgag	gcagcgcggc	tatcgtggct	3720
ggccacgacg	ggcgttcctt	gcgcagctgt	gctcgacgtt	gtcactgaag	cgggaaggga	3780
ctgggtgcta	ttgggcgaag	tgccggggca	ggatctcctg	tcactctacc	ttgctcctgc	3840
cgagaaagta	tccatcatgg	ctgatgcaat	gcggcggctg	catacgcttg	atccggctac	3900
ctgcccattc	gaccaccaag	cgaaacatcg	catcgagcga	gcacgtactc	ggatggaagc	3960
cggctcttgt	gatcaggatg	atctggacga	agagcatcag	gggctcgcgc	cagccgaact	4020
gttcgccagg	ctcaaggcgc	gcatgcccga	cgtcgaggat	ctcgtcgtga	cccatggcga	4080
tgccctgcttg	ccgaatatca	tgggtgaaaa	tggccgcctt	tctggattca	tcgactgtgg	4140
ccggctgggt	gtggcggacc	gctatcagga	catagcgttg	gctaccctg	atatgtctga	4200
agagcttggc	ggcgaatggg	ctgaccgctt	cctcgtgctt	tacggtatcg	ccgctcccga	4260
ttcgcagcgc	atcgccctct	atcgccctct	tgacgagttc	ttctgagcgg	gactctgggg	4320
ttcgaaatga	ccgaccaagc	gacgcccac	ctgccaatcac	gatggccgca	ataaaatatc	4380
tttattttca	ttacatctgt	gtgttggttt	tttgtgtgaa	tcgatagcga	taaggatccg	4440
cgtatggtgc	actctcagta	caatctgctc	tgatgccgca	tagttaagcc	agccccgaca	4500
cccgccaaaca	cccgttgacg	cgccctgacg	ggcttgtctg	ctcccggcat	ccgcttacag	4560
acaagctgtg	accgtctccg	ggagctgcag	gtgtcagagg	ttttcaccgt	catcacccga	4620
acgcgcgaga	cgaaagggcc	tcgtgatacg	cctattttta	taggttaatg	tcattgataat	4680
aatgggtttct	tagacgtcag	gtggcacttt	tcggggaaat	gtgcgcggaa	cccctatttg	4740
tttatttttc	ttaatacat	caaataatgt	tccgctcatg	agacaataac	cctgataaat	4800
gcttcaataa	tattgaaaaa	ggaagagtat	gagtattcaa	catttccgtg	tcgcccttat	4860
tccttttttt	gcggcatttt	gccttctctg	ttttgtctcac	ccagaaaacgc	tggtgaaaagt	4920
aaaagatgct	gaagatcagt	tgggtgcacg	atcgaaactg	atcgaaactg	atctcaacag	4980
cggttaagatc	cttgagagtt	ttcgccccga	agaacgtttt	ccaatgatga	gcacttttaa	5040
agttctgcta	tgtggcgctg	tattatcccg	tattgacgcc	gggcaagagc	aaactcggctg	5100
ccgcatacac	tattctcaga	atgacttgg	tgagtactca	ccagtcacag	aaaagcatct	5160
tacggatggc	atgacagtaa	gagaattatg	cagtgtctgcc	ataaccatga	gtgataaac	5220
tgccggccaac	ttacttctga	caacgatcgg	aggaccgaag	gagctaaccg	cttttttgca	5280
caacatgggg	gatcatgtaa	ctcgccctga	tcgttgggaa	ccggagctga	atgaagccat	5340
accaaaagc	gagcgtgaca	ccacgatgcc	tgtagcaatg	gcaacaacgt	tgcgcaaaact	5400
attaactggc	gaactactta	ctctagcttc	ccggcaacaa	ttaatagact	ggatggaggc	5460
ggataaaagt	gcaggaccac	ttctgcgctc	ggcccttccg	gctggctgg	ttattgctga	5520
taaatctgga	gccggtgagc	gtgggtctcg	gggtatcatt	gcagcactgg	ggccagatgg	5580
taagccctcc	cgtatcgtag	ttatctacac	gacggggagt	cagggaacta	tggtatgaacg	5640
aaatagacag	atcgctgaga	taggtgcctc	actgattaag	cattggtaac	tgtcagacca	5700
agtttactca	tataactttt	agattgattt	aaaacttcat	ttttaattta	aaaggatcta	5760
ggtgaagatc	cttttttgata	atctcatgac	caaaatccct	taacgtgagt	tttcgttcca	5820
ctgagcgtca	gaccccgtag	aaaagatcaa	aggatcttct	tgagatcctt	tttttctgcg	5880
cgtaatctgc	tgcttgcaaa	caaaaaaacc	accgctacca	gcggtgggtt	gtttgcccga	5940
tcaagagcta	ccaactcttt	ttccgaaggt	aaactggctt	agcagagcgc	agataccaaa	6000
tactgtcctt	ctagtgtagc	cgtagttagg	ccaccacttc	aagaactctg	tagcacccgc	6060
tacataacct	gctctgctaa	tcctgttacc	agtggctgct	gccagtggcg	ataagtcgtg	6120
tcttaccggg	ttggactcaa	gacgatagtt	accggataag	gcgcagcgg	cgggctgaac	6180
gggggggttc	tgcacacagc	ccagcttgga	gcgaacgacc	tacaccgaac	tgagatacct	6240
acagcgtgag	ctatgagaaa	gcgccacgct	tccgaagg	agaaaggcgg	acaggtatcc	6300
ggtaagcggc	agggtcggaa	caggagagcg	cacgagggag	cttccagggg	gaaacgcctg	6360
gtatctttat	agtcctgtcg	ggtttcgcca	cctctgactt	gagcgtcgat	ttttgtgatg	6420
ctcgtcaggg	gggcggagcc	tatggaaaaa	cgccagcaac	gcggcctttt	tacggttcct	6480
ggccttttgc	tggccttttg	ctcacatggc	tcgacagatc	t		6521



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/17452

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/04; C12N 15/85,87,90; A01N 43/34; C12N 15/09

US CL : 536/23.1; 435/320.1,325,419,455,467; 514/44; 800/21

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1; 435/320.1,325,419,455,467; 514/44; 800/21

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
APS (EAST); STN (MEDLINE, BIOSIS, CAPLUS)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 01/07572 A2 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 01 February 2001 (01.02.2001), see entire document.	1,2,7,8,11,23-25,27,50
X --- Y	WO 00/11155 A1 (THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY) 02 March 2000 (02.03.2000), see entire document.	3-6,9,10,12,13,26,28-1,2,7,8,11,23-25,27,50
X --- Y	US 6,171,861 B1 (HARTLEY et al.) 09 January 2001 (09.01.2001), see entire document, especially sequence listing.	3-6,9,10,12,13,26,28-1-3,8,10-12,23-27,33,36,39,40,43-45,50,54,55,85 4-7,9,13-22,28-30,32,34,35,37,38,41,42,46-49,51-53,56,67-71,84,86



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

01 November 2002 (01.11.2002)

Date of mailing of the international search report

27 NOV 2002

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Daniel M Sullivan

Telephone No. 703-308-0196



# INTERNATIONAL SEARCH REPORT

PCT/US02/17452

## C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 96/40724 A1 (LIFE TECHNOLOGIES, INC.) 19 December 1996 (19.12.96), see entire document, especially SEQ ID NO:1-16.	1-3,8,10-12,23-27,33,36,39,40,43-45,50,54,55,85 ----- 1-3,8,10-12,23-27,33,36,39,40,43-45,50,54,55,85
Y	WO 94/00569 A1 (GENPHARM INTERNATIONAL, INC.) 06 January 1994 (06.01.94), see entire document.	72-78
Y	WO 94/23049 A2 (THE JOHNS HOPKINS UNIVERSITY) 13 October 1994 (13.10.94), see entire document.	72-78
X --- Y	US 5,721,118 A (SCHEFFLER) 24 February 1998 (24.02.98), see entire document, especially Example 3.	14,15,22-25,57 ----- 16-21,26,58-64,72-78,84,85,98-105,123
X --- Y	US 5,948,653 A (PATI et al.) 07 September 1999 (07.09.99), see entire document, especially the first full paragraph in column 29.	14,15,22-25,57 ----- 16-21,26,58-64,72-78,84,85,98-105,123
X --- Y	GB 2 331 752 A (MEDICAL RESEARCH COUNCIL) 02 June 1999 (02.06.99), see entire document, especially pages 8-10.	109,123 ----- 72-78,110-122
A	HALDIMANN et al. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. J Bacteriol. November 2001, Vol. 183 No. 21, pages 6384-6393, see entire document.	65,66
X,P --- Y,P	MORALLI et al. Insertion of a loxP site in a size-reduced human accessory chromosome. Cytogenet. Cell Genet. 2001, Vol. 94 No. 3-4, pages 113-120, see entire document.	91-93,106,108 ----- 94-97,107
X --- Y	LORBACH et al. Site-specific recombination in human cells catalyzed by phage lambda integrase mutants. J. Mol. Biol. March 2000, Vol. 296 No. 5, pages 1175-1181, see entire document.	1,2,8,50 ----- 3-7,9-13,23-32,51-56,84,85
X --- Y	CALL et al. A cre-lox recombination system for the targeted integration of circular yeast artificial chromosomes into embryonic stem cells. Hum. Mol. Genet. July 2000, Vol. 9, No. 12, pages 1745-1751, see entire document.	14,15,22-26,57,59 ----- 16-21,58,60,84,85,98-105
Y,P	HADLACZKY, G. Satellite DNA-based artificial chromosomes for use in gene therapy. Curr. Opin. Mol. Ther. April 2001, Vol. 3 No. 2, pages 125-132, see entire document.	1-123

## INTERNATIONAL SEARCH REPORT

PCT/US02/17452

### BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-64, 67-71, 79, 84-86, 91-108, and 123, drawn to eukaryotic recombinogenic chromosomes used for introducing heterologous nucleic acids into a chromosome and the resulting cells.

Group II, claim(s) 65-66 and 87-88, drawn to a lambda intR mutain.

Group III, claim(s) 72-78, drawn to the production of transgenic animals.

Group IV, claim(s) 80, drawn to the production of an artificial chromosome library.

Group V, claim(s) 81-83, drawn to a library of cells for genomic screening.

Group VI, claim(s) 89-90, drawn to a modified iron-induced promoter.

Group VII, claim(s) 109-122, drawn to a method for screening compounds and their effects on regulatory regions.

and it considers that the International Application does not comply with the requirements of unity of invention (Rules 13.1, 13.2 and 13.3) for the reasons indicated below:

The inventions listed as Groups I-VII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical feature of Group I which defines an advance over the art is a eukaryotic chromosome containing recombinogenic sites that can be used to introduce heterologous nucleic acids into chromosomes, and cells containing these recombinogenic chromosomes.

The special technical feature of Group II involves a lambda intR mutain. This feature defines an advance over Group I in that it involves a protein that is not required for the technical features as set forth above in Group I.

The special technical feature of Group III involves the production of a transgenic animal, which represents a second method for using the invention as set forth above in Group I.

The special technical feature of Group IV involves the production of artificial chromosome expression system libraries, which represents a third method for using the invention as set forth above in Group I.

The special technical feature of Group V involves a library of cells containing the artificial chromosome expression system libraries set forth above in Group IV, and represents a first product resulting from Group IV.

The special technical feature of Group VI involves a modified iron-inducible promoter. This feature defines an advance over Group I in that it involves a promoter that is not required for the technical features as set forth above in Group I.

The special technical feature of Group VII involves a method for screening compounds for their effect on regulatory regions, which represents a fourth method of using the invention as set forth above in Group I.

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US02/17452

## Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:  
Please See Continuation Sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐  
☐

The additional search fees were accompanied by the applicant's protest.  
No protest accompanied the payment of additional search fees.